Biostatistică și Epidemiologie

\ Primer pentru profesioniști în sănătate și biomedicali

Ediția a treia

Sylvia Wassertheil-Smoller

Biostatistică și Epidemiologie

Ediția a treia

Springer
New York
Berlin
Heidelberg
Hong Kong
Londra
Milano
Paris
Tokyo

Sylvia Wassertheil-Smoller

Profesor și șef, Divizia de Epidemiologie,
Departamentul de Epidemiologie și Sănătatea Populației
Titular al Fundației Dorothy Maneleoff
și Catedră Molly Rosen de Medicină Socială
Colegiul de Medicină Albert Einstein

Biostatistică și Epidemiologie

Un primer pentru profesioniștii din sănătate și biomedicali

Ediția a treia

Cu 22 de ilustrații

Springer

Sylvial Wassertheil-Smoller
Departamentul de Epidemiologie și Sănătatea Populației Divizia de Epidemiologie
Colegiul de Medicină Albert Einstein
Bronx, NY 10461-1602
STATELE UNITE ALE AMERICII
smoller@aecom.yu.edu

Datele de catalogare în publicație ale Bibliotecii Congresului
Wassertheil-Smoller, Sylvia
Biostatistică și epidemiologie: un primer pentru sănătate și profesioniști biomedicali / de Sylvia Wassertheil-Smoller.—ed. a III-a.
p. cm.
Include referințe bibliografice și index.
ISBN 0-387-40292-6 (hârtie alk.)
1. Epidemiologie—Metode statistice. 2. Studii clinice—Metode statistice. I. Titlu.
RA652.2.M3W37 2003
614.4'072—dc22 2003058444

Imprimat pe hârtie fără acid.

ISBN 0-387-40292-6 Tipărit pe hârtie fără acid.

© 2004, 1995, 1990 Springer-Verlag New York, Inc.
Toate drepturile rezervate. Această lucrare nu poate fi tradusă sau copiată integral sau parțial fără - permisiunea scrisă a editorului (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, SUA), cu excepția unor scurte fragmente în legătură cu recenzii sau analize academice. Este interzisă utilizarea în legătură cu orice formă de stocare și regăsire a informațiilor, adaptare electronică, software de calculator sau prin metodologie similară sau diferită cunoscută acum sau dezvoltată în continuare.
Utilizarea în această publicație a numelor comerciale, mărcilor comerciale, mărcilor de servicii și termenilor similari, chiar dacă nu sunt identificați ca atare, nu trebuie să fie considerată o expresie a opiniei cu privire la faptul dacă sunt sau nu supuse drepturilor de proprietate.

Tipărit în Statele Unite ale Americii. **(BPR/MVY) {NECESIT?}**
987654321 SPIN 10935774
Springer-Verlag face parte din *Springer Science+Business Media springeronline.com*

Spre Iordania

PREFAȚĂ LA EDIȚIA A TREIA

și rezumatul prefețelor la primele două ediții

Această carte, prin numeroasele sale ediții, a continuat să se adapteze la domeniile în evoluție de cercetare în epidemiologie și statistică, menținând în același timp obiectivul inițial de a fi neamenințătoare, de înțeles și accesibilă celor cu cunoștințe limitate sau fără cunoștințe în matematică. Două domenii noi sunt acoperite în a treia ediție: epidemiologia genetică și etica cercetării.

Odată cu secvențierea genomului uman, a existat o evoluție a cercetării asupra bazei genetice a sănătății și bolii și, în special, a interacțiunilor dintre gene și expunerile la mediu. Literatura medicală în epidemiologia genetică este în expansiune vastă și este necesară o anumită cunoaștere a modelelor epidemiologice și o cunoaștere a metodelor statistice utilizate în astfel de cercetări pentru a putea aprecia noile descoperiri. Astfel, această ediție include un nou capitol despre epidemiologia genetică, precum și un apendice care descrie elementele de bază necesare pentru înțelegerea cercetării genetice. Un astfel de material nu se găsește de obicei în cărțile de epidemiologie sau statistică de prim nivel, dar este prezentat aici într-un mod de bază și, sperăm, ușor de înțeles, pentru cei care nu sunt familiarizați cu domeniul. Al doilea capitol nou se referă la etica cercetării, de asemenea, care nu este acoperit de obicei în manualele de bază, dar este extrem de important în toate cercetările umane. A fost adăugat și material nou la mai multe capitole existente.

Obiectivele principale ale primei ediții încă se aplică. Prezentarea - materialului este menită să ofere o înțelegere a principiilor care stau la baza, precum și ghiduri practice despre „cum să o faci" și „cum să o interpretezi". Subiectele incluse sunt cele care sunt cele mai frecvent utilizate sau la care se face referire în literatură. Există câteva caracteristici de reținut care pot ajuta cititorul în utilizarea acestei cărți:

(a) Cartea începe cu o discuție despre filozofia și logica științei și principiile care stau la baza testării a ceea ce credem față de realitatea experiențelor noastre. Deși o astfel de discuție, în sine, nu va ajuta cititorul

să „facă un test t”, cred că este important să oferim o introducere în cadrul de bază al domeniului epidemiologiei și statisticii, pentru a înțelege de ce facem ceea ce facem.

(b) Multe dintre subsecțiuni sunt singure; adică cititorul poate apela la subiectul care îl interesează și poate citi materialul în ordine secvențială. Astfel, cartea poate fi folosită de cei care au nevoie de ea în scopuri speciale. Cititorul este liber să sări peste acele subiecte care nu sunt de interes fără a fi prea mult împiedicat în continuarea lecturii. Ca urmare, există o oarecare redundanță. Totuși, în experiența mea de predare, am constatat că este mai bine să greșesc din partea redundanței decât din partea lipsei.

(c) Referințele încrucișate la alte secțiuni relevante sunt incluse atunci când sunt necesare explicații suplimentare. Atunci când dezvoltarea unui subiect depășește sfera acestui text, cititorul este îndrumat către alte cărți care tratează materialul în profunzime sau la un nivel matematic superior. O listă de texte recomandate este oferită aproape de sfârșitul cărții.

(d) Anexele oferă exemple de calcule pentru diverse statistici descrise în text. Acest lucru face o citire mai ușoară a textului, oferind în același timp cititorului instrucțiuni mai specifice despre cum să efectueze efectiv unele dintre calcule.

Scopurile celei de-a doua ediții sunt păstrate și în această a treia ediție . Cea de-a doua ediție a luat naștere din feedback-ul studenților care au - indicat că au apreciat claritatea și concentrarea asupra subiectelor legate în mod specific de munca lor. Cu toate acestea, unii utilizatori au ratat acoperirea mai multor subiecte importante. În consecință, au fost adăugate secțiuni pentru a include un capitol complet despre măsurile calității vieții și diverse scale psihologice, care sunt din ce în ce mai utilizate în studiile clinice; o extindere a capitolului despre probabilitate, cu introducerea mai multor metode neparametrice; clarificarea unor concepte care au fost abordate mai concis în prima ediție; și adăugarea mai multor apendice (care oferă eșantion de calcule ale testului exact al lui Fisher, Kruskal-Wallis

test și diverși indici de fiabilitate și de receptivitate a scalelor utilizate în măsurile calității vieții).

Este nevoie de un echilibru delicat pentru a menține cartea concisă și de bază și totuși pentru a o face suficient de incluzivă pentru a fi utilă unui public larg. Sper că această carte va fi utilă diverselor grupuri de oameni din domeniul sănătății, precum și celor din domenii conexe. Materialul este destinat (1) medicilor care fac cercetări clinice, precum și celor care fac cercetări de bază; (2) pentru studenți - medicină, facultate și absolvenți; (3) pentru personalul de cercetare în diferite calități; și (4) pentru oricine este interesat de logica și metodologia biostatisticii și epidemiologiei. Principiile și metodele descrise aici sunt aplicabile în diferite domenii de fond, inclusiv medicină, sănătate publică, psihologie și educație. Desigur, nu toate subiectele care sunt relevante în mod specific pentru fiecare dintre aceste discipline pot fi acoperite în acest scurt text.

Bronx, New York Sylvia Wassertheil-Smoller

MULȚUMIRI

Vreau să-mi exprim recunoştinţa pentru învăţătura inspirată a doctorului Jacob Cohen, acum decedat, care m-a pornit pe această cale; lui Jean Almond, care mi-a făcut personal posibil să continui; și colegilor și studenților mei de la Colegiul de Medicină Albert Einstein, care îl fac distractiv.

Aprecierea mea se îndreaptă către cei care au criticat prima și a doua ediție: Dr. Brenda Breuer, Dr. Ruth Hyman, Dr. Ruth Macklin, Dr. Dara Lee, Dr. CJ Chang și Dr. Jordan Smoller, precum și numeroșii studenți care mi-au oferit feedback. Mulțumesc Dr. Gloria Ho, Dr. Charles Hall, Dr. Charles Kooperberg, Mimi Goodwin și Dr. Paul Bray pentru ajutor în editarea noului material din a treia ediție și pentru sugestiile lor utile.

Mulțumiri foarte speciale fiului meu, dr. Jordan Smoller, pentru - ajutorul valoros oferit cu materialul despre genetică, editarea amănunțită a cărții și pentru orice altceva, întotdeauna.

Îi sunt foarte îndatorat doamnei Ann Marie McCauley, pentru munca sa pricepută și remarcabilă la edițiile anterioare și doamnei Darwin Tracy pentru pregătirea artistică a noului manuscris, răbdarea , angajamentul și munca grea remarcabile pentru această carte.

În cele din urmă, dragostea și recunoștința mea profundă sunt către soțul meu, Walter Austerer, pentru ajutorul, încurajarea și răbdarea lui.

Bronx, New York Sylvia Wassertheil-Smoller

CUPRINS

PREFAȚĂ LA EDIȚIA A TREIA Vi
MULȚUMIRI ix

CAPITOLUL 1. METODA ŞTIINŢIFICĂ 1
1.1 Logica raționamentului științific 1
1.2 Variabilitatea fenomenelor necesită o analiză statistică 6
1.3 Inferență inductivă: Statistica ca tehnologie a metodei științifice 7
1.4 Proiectarea studiilor 8
1.5 Cum se cuantifică variabilele 10
1.6 Ipoteza nulă 11
1.7 De ce testăm ipoteza nulă? 12
1.8 Tipuri de erori 14
1.9 Nivel de semnificație și tipuri de eroare 15
1.10 Consecințele erorilor de tip I și de tip II 16

CAPITOLUL 2. PUȚINĂ DE PROBABILITATE 19
2.1 Ce este probabilitatea? 19
2.2 Combinarea probabilităților 20
2.3 Probabilitate condiționată 23
2.4 Probabilitatea Bayesiană 24
2.5 Cote și probabilitate 25
2.6 Raportul de probabilitate 26
2.7 Rezumatul probabilității 27

CAPITOLUL 3. CELE MAI MULT DESPRE STATISTICĂ 29
3.1 Chi-Pătrat pentru 2 x 2 mese 29
3.2 Testul McNemar 34
3.3 Kappa 35
3.4 Descrierea unei populații: Utilizarea standardului Abaterea 36
3.5 Semnificația abaterii standard: Normal Distributie 40
3.6 Diferența dintre abaterea standard și Eroare standard 42

3.7 Eroare standard a diferenței dintre două mijloace 46
3.8 Scorurile Z și distribuția normală standardizată 47
3.9 Statistica t 51
3.10 Valori eșantioane și valori ale populației revizuite 52
3.11 O întrebare de încredere 53
3.12 Limite de încredere și intervale de încredere 55
3.13 Grade de libertate 56
3.14 Intervale de încredere pentru proporții 57
3.15 Intervale de încredere în jurul diferenței dintre Două mijloace 58
3.16 Comparații între două grupuri 60
3.17 Testul Z pentru compararea a două proporții 60
3.18 t-Test pentru diferența dintre mijloacele a doi Grupuri independente: principii 62
3.19 Cum se face un test t: un exemplu 64
3.20 Perechea potrivită t-Test 66
3.21 Când să nu faci o mulțime de teste t: problema de Teste multiple de semnificație 67
3.22 Analiza variației: comparație între mai multe Grupele 69
3.23 Principii 69
3.24 Procedura Bonferroni: O abordare a realizarii Comparații multiple 72
3.25 Analiza variației atunci când sunt două independente Variabile: ANOVA cu doi factori 74
3.26 Interacțiunea între două variabile independente 75
3.27 Exemplu de ANOVA bidirecțional 76
3.28 Testul Kruskal-Wallis pentru a compara mai multe grupuri 77
3.29 Asociere și cauza: coeficientul de corelație 78
3.30 Cât de sus este mare? 79
3.31 Căi cauzale 79
3.32 Regresia 82
3.33 Legătura dintre regresia liniară și Coeficientul de corelație 84
3.34 Regresia liniară multiplă 84
3.35 Rezumat până acum 86

CAPITOLUL 4. CELE MAI MULT DESPRE EPIDEMIOLOGIE 87
4.1 Utilizările epidemiologiei 87
4.2 Câteva concepte epidemiologice: ratele mortalității 88
4.3 Tarife ajustate în funcție de vârstă 90
4.4 Ratele de incidență și prevalență 92
4.5 Rata de mortalitate standardizată 94
4.6 Persoană-ani de observație 94
4.7 Variabile dependente și independente 96
4.8 Tipuri de studii 96
4.9 Analize transversale versus longitudinale ale datelor 97
4.10 Măsuri ale riscului relativ: Deduceri din Studii prospective: studiul Framingham 101
4.11 Calculul riscului relativ din studii prospective 103
4.12 Odds Ratio: Estimarea riscului relativ din Studii de caz-control 104
4.13 Risc atribuibil 107
4.14 Prejudecata de răspuns 109
4.15 Variabile de confuzie 111
4.16 Potrivire 112
4.17 Regresia logistică multiplă 113
4.18 Confuzie prin indicația 116
4.19 Analiza supraviețuirii: metodele tabelului de viață 117
4.20 Cox Proportional Hazards Model 120
4.21 Selectarea variabilelor pentru modelele multivariate 122
4.22 Interacțiuni: modele aditive și multiplicative 124
Rezumat: 127
CAPITOLUL 5. MAI MULTE DESPRE PROIECTARE 129
5.1 Sensibilitate, specificitate și concepte înrudite 129
5.2 Punctul de limită și efectele acestuia asupra sensibilității și specificității 136

CAPITOLUL 6. CELE MAI MULT DESPRE TRESTELE CLINICE 141
6.1 Caracteristicile studiilor clinice randomizate 141
6.2 Scopurile randomizării 143
6.3 Cum se efectuează o atribuire aleatorie 144
6.4 Teste cu două cozi versus testul cu o singură coadă 145

6.5 Studiul clinic ca „Standard de aur” 146
6.6 Regresia spre medie 147
6.7 Analiza intenției de tratare 150
6.8 Cât de mare ar trebui să fie studiul clinic? 151
6.9 Ce este implicat în calculul dimensiunii eșantionului? 153
6.10 Cum se calculează dimensiunea eșantionului pentru diferență Între două proporții 157
6.11 Cum se calculează dimensiunea eșantionului pentru a testa diferența Între două mijloace 158

CAPITOLUL 7. CELE MAI MULT DESPRE CALITATEA VIEI 161
7.1 Construcție la scară 162
7.2 Fiabilitate 162
7.3 Valabilitate 164
7.4 Reactivitate 165
7.5 Câteva capcane potențiale 167

CAPITOLUL 8. CELE MAI MULT DESPRE EPIDEMIOLOGIA GENETICĂ 171
8.1 O nouă eră științifică 171
8.2 Prezentare generală a epidemiologiei genetice 172
8.3 Studii pe gemeni 173
8.4 Studii de legături și asocieri 175
8.5 Scor LOD: Statistica de legătură 178
8.6 Studii de asociere 179
8.7 Teste de dezechilibru al transmisiei (TDT) 181
8.8 Câteva concepte și complexități suplimentare de studii genetice 185
CAPITOLUL 9. ETICA ŞI STATISTICA CERCETĂRII 189
9.1 Ce legătură are statisticile cu asta? 189
9.2 Protecția subiectelor de cercetare umane 190
9.3 Consimțământ informat 192
9.4 Echilibrul 194
9.5 Integritatea cercetării 194
9.6 Politicile de autor 195
9.7 Comisiile de monitorizare a datelor și a siguranței 196

9.8 Rezumatul 196

Postscript CÂTEVA COMENTARII DE DESPĂZIRE LA IMPACTUL EPIDEMIOLOGIEI ASUPRA OMULUI VIEȚI 197

Anexa A. VALORI CRITICE ALE CHI-PĂTRATUL, Z, ȘI t 199

TESTUL EXACT AL FISHER 201

Anexa C. KRUSKAL-WALLIS NEPARAMETRIC TEST PENTRU COMPARAREA MAI MULTE GRUPE 203

Anexa D. CUM SE CALCULEAZĂ O CORELATIE COEFICIENTUL 205

Anexa E. AJUSTAREA Vârsta 207

Anexa F. LIMITE DE ÎNCREDERE PRIVIND RAPORTELE DE COTE 211

Anexa G. RELATIE IN FORMA DE „J” SAU „U”. ÎNTRE DOUĂ VARIABILE 213

Anexa H. DETERMINAREA ADECVĂRII Schimbați scorurile 217

Anexa I. PRINCIPII GENETICE 221

REFERINȚE 227
LECTURI SUGERATE 233
INDEX 237

Capitolul 1
METODA ŞTIINŢIFICĂ

Știința este construită cu fapte, așa cum o casă este cu pietre. Dar o colecție de fapte nu este mai mult o știință decât un morman de pietre este o casă.

Jules Henri Poincare La știința și ipoteza (1908)

1.1 Logica raționamentului științific

Scopul științei este de a descoperi „adevărul". Cum să decidem că ceva este adevărat? Avem două instrumente la dispoziție pentru a continua cercetarea științifică:

Avem simțurile noastre, prin care experimentăm lumea și facem *observații.*

Avem capacitatea de a raționa, ceea ce ne permite să facem *inferențe logice.*

În știință impunem *logica* acestor observații.

În mod clar, avem nevoie de ambele instrumente. Toată logica din lume nu va crea o observație, iar toate observațiile individuale din lume nu vor crea în sine o teorie. Există două tipuri de relații între mintea științifică și lume, două tipuri de logică pe care le impunem – *deductivă și inductivă,* așa cum este ilustrat în Figura 1.1.

În *inferența deductivă* , deținem o teorie și pe baza ei facem o predicție a consecințelor acesteia. Adică, prezicem care ar trebui să fie observațiile. De exemplu, putem susține o teorie a învățării care spune că întărirea pozitivă are ca rezultat o învățare mai bună decât pedeapsa, adică recompensele funcționează mai bine decât pedepsele. Din această teorie prezicem că studenții la matematică care sunt lăudați pentru dreptul lor

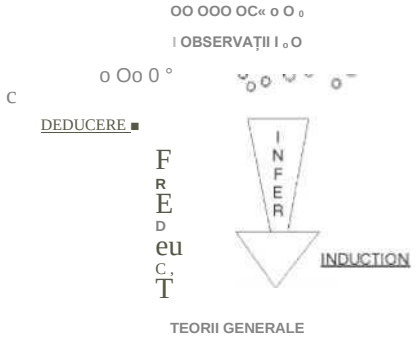


Figura 1.1

răspunsurile din timpul anului se vor descurca mai bine la examenul final decât cei care sunt pedepsiți pentru răspunsurile lor greșite. Trecem de la general – teorie – la specific – observații. Aceasta este cunoscută ca metoda hipotetico-deductivă.

În *inferența inductivă,* trecem de la specific la general. Facem multe observații, discernem un model, facem o generalizare și deducem o explicație. De exemplu, la Spitalul General din Viena, în anii 1840, s-a observat că femeile care nasc mureau cu o rată ridicată de febră puerperală, o generalizare care a provocat teroare la viitoarele mame. Un tânăr medic pe nume Ignaz Phillip Semmelweis a legat observația că studenții la medicină care efectuează examinări vaginale au făcut acest lucru imediat după ce au venit din camera de disecție, spălându-se rar pe mâini între ele, cu observația că un coleg care și-a tăiat accidental degetul în timp ce diseca un cadavru a murit de o boală exact ca cea care ucidea mamele. El a dedus explicația că cauza morții a fost introducerea de material cadavru într-o rană. Consecința practică a acelui salt creativ al imaginației a fost eliminarea febrei puerperale ca flagel al nașterii prin solicitarea ca medicii să se spele pe mâini înainte de a face o naștere! Abilitatea de a face astfel de salturi creative din generalizări este produsul minților științifice creative.

În general, s-a considerat că epidemiologii folosesc inferența inductivă. De exemplu, cu câteva decenii în urmă s-a observat că femeile păreau să facă atacuri de cord cu aproximativ 10 ani mai târziu decât bărbații. Un salt creativ al imaginației a dus la concluzia că hormonii femeilor au fost cei care le-au protejat până la menopauză. EUREKA! Ei au dedus că, dacă

estrogenul este bun pentru femei, trebuie să fie bun pentru bărbați și au prezis că observațiile ar corobora această deducere. A fost întreprins un studiu clinic care a dat bărbaților cu risc crescut de apariție a estrogenului în doze destul de mari, 2,5 mg pe zi sau aproximativ de patru ori doza utilizată în prezent la femeile aflate în postmenopauză. În mod deloc surprinzător, bărbații nu au apreciat efectele secundare, dar, în mod surprinzător pentru cercetători , bărbații din grupul cu estrogeni au avut rate mai mari de boli coronariene și mortalitate decât cei tratați cu placebo. [2] Ceea ce a fost bun pentru gâscă s-ar putea să nu fie atât de bun pentru gâscă. Studiul a fost întrerupt și estrogenul ca măsură preventivă a fost abandonat timp de câteva decenii.

În acest timp, multe studii observaționale prospective au indicat că înlocuirea estrogenului administrată femeilor aflate în postmenopauză a redus riscul de boli de inimă cu 30-50%. Aceste observații au condus la concluzia inductivă că înlocuirea hormonală în postmenopauză este protectoare, adică observațiile au condus la teorie. Cu toate acestea, această teorie trebuie testată în studii clinice. Primul astfel de studiu de substituție hormonală la femeile care aveau deja boli de inimă, Heart and Estrogen/Progestin Replacement Study (HERS) nu a găsit nicio diferență în ratele bolilor de inimă între grupul de tratament activ și grupul placebo, dar a găsit o creștere precoce a evenimentelor de boli de inimă în primul an de studiu și un beneficiu ulterior al hormonilor după aproximativ 2 ani. Deoarece acesta a fost un studiu pe femei cu boală cardiacă stabilită, a fost un studiu secundar de prevenire și nu răspunde la întrebarea dacă femeile fără boli de inimă cunoscute ar beneficia de înlocuirea hormonală pe termen lung . Această întrebare a fost abordată de Inițiativa pentru sănătatea femeilor (WHI), care este descrisă într-o secțiune ulterioară.

Scopul exemplului este de a ilustra modul în care observațiile (că femeile suferă de boli de inimă mai târziu decât bărbații) duc la o teorie (că hormonii sunt protectori), care prezice noi observații (că vor exista mai puține atacuri de cord și decese în rândul celor pe hormoni), care ar putea întări teoria, până când este testată într-un studiu clinic care poate fie să o coroboreze și să o ducă la o nouă testare, fie să o înlăture. dacă prezice mai bine observaţii noi. Deci, există o interacțiune constantă între inferența inductivă (bazată pe observații) și inferența deductivă (bazată pe teorie), până când ne apropiem din ce în ce mai mult de „adevăr”.

Cu toate acestea, există un alt punct în această poveste. Teoriile nu ies

doar din fapte. Trebuie să existe un substrat din care să iasă teoria. Poate că acel substrat este o altă teorie anterioară care s-a dovedit a fi inadecvată pentru a explica aceste noi observații și că teoria, la rândul său, a înlocuit o teorie anterioară. În orice caz, unul din „substratul" este „mintea pregătită" a investigatorului. Dacă investigatorul este cardiolog, de exemplu, el sau ea este instruit să privească fenomenele medicale din perspectiva cardiologiei și cunoaște teoriile precedente și punctele forte și defectele acestora. Dacă diologul auto nu ar fi avut o astfel de pregătire, el sau ea ar fi putut să nu fi văzut legătura. Sau, cu o pregătire diferită, investigatorul ar putea sări la o inferență cu totul diferită. Epidemiologul trebuie să lucreze într-o echipă interdisciplinară pentru a aduce diverse perspective asupra unei probleme și pentru a înrola minți „pregătite" în moduri diferite.

Întrebarea este, cât de bine rezistă o teorie în fața noilor observații? Atunci când multe studii oferă dovezi afirmative în favoarea unei teorii, ne crește asta credința în ea? Dovezile afirmative înseamnă mai multe exemple care sunt în concordanță cu teoria. Dar în ce măsură dovezile susținătoare întăresc o afirmație? Cei care cred că inducția este logica adecvată a științei susțin opinia că dovezile afirmative sunt ceea ce întărește o teorie.

O altă abordare este cea a lui Karl Popper, poate unul dintre cei mai importanți teoreticieni ai științei. Popper susține că inducția rezultată din acumularea de dovezi afirmative nu întărește o teorie. Inducția, la urma urmei, se bazează pe credința noastră că lucrurile neobservate vor fi ca cele observate sau că viitorul va fi ca trecutul. De exemplu, vedem o mulțime de lebede albe și facem afirmația că toate lebedele sunt albe. Această afirmație este susținută de multe observații. De fiecare dată când vedem o altă lebădă albă, avem mai multe dovezi de susținere. Dar nu putem dovedi că toate lebedele sunt albe, indiferent câte lebede albe vedem.

Pe de altă parte, această afirmație poate fi doborâtă de vederea unei singure lebede negre. Acum ar trebui să ne schimbăm afirmația să spunem că majoritatea lebedelor sunt albe și că există unele lebede negre. Această afirmație este probabil mai aproape de adevăr. Cu alte cuvinte, putem respinge afirmația cu un singur exemplu, dar nu o putem dovedi cu multe. (Aserțiunea că toate lebedele sunt albe este mai degrabă o generalizare descriptivă decât o teorie. O teorie are un înțeles mai bogat care încorporează explicații cauzale și mecanisme care stau la baza. Afirmațiile, precum cele referitoare la culoarea lebedelor, pot fi componente ale unei teorii.)

Potrivit lui Popper, metodologia potrivită este să postulezi o teorie, sau o conjectură, așa cum o numește el, și să încerci să demonstrezi că este falsă. Cu cât supraviețuiește mai multe astfel de încercări de distrugere, cu atât mai puternică este dovezile pentru aceasta . Scopul este de a concepe încercări din ce în ce mai agresive de a dărâma afirmația și de a vedea dacă încă supraviețuiește. Dacă nu supraviețuiește unei tentative de *falsificare,* atunci teoria este renunțată și înlocuită cu alta. El numește aceasta metoda *presupunerilor și infirmărilor.* Înaintarea științei către „adevăr" are loc prin eliminarea teoriilor ale căror predicții nu sunt confirmate de observații sau teorii care nu sunt testabile în totalitate, mai degrabă decât prin susținerea teoriilor cu mai multe exemple de unde funcționează. Teoriile *științifice utile sunt potențial falsificabile.*

Teoriile netestabile sunt acelea în care o varietate de observații contradictorii ar putea fi fiecare în concordanță cu teoria. De exemplu, luați în considerare teoria psihanalitică a lui Freud. Teoria complexului lui Oedip spune că un copil este îndrăgostit de părintele de sex opus. Un băiat își dorește mama și vrea să-și distrugă tatăl. Dacă observăm că un bărbat spune că își iubește mama, asta se potrivește cu teoria. Dacă observăm că un bărbat spune că își urăște mama, asta se potrivește și cu teoria, care ar spune că „formarea reacției" este cea care îl face să-și nege adevăratele sentimente . Cu alte cuvinte, indiferent ce spune omul, nu ar putea falsifica teoria pentru că ar putea fi explicată prin ea. Deoarece nicio observație nu ar putea falsifica teoria lui Oedip, poziția sa ca teorie științifică ar putea fi pusă sub semnul întrebării.

O a treia opinie, și cea mai rezonabilă, este că progresul științei necesită atât inferențe inductive, cât și deductive. Un anumit punct de vedere oferă un cadru pentru observații, care duc la o teorie care prezice noi observații care modifică teoria, ceea ce conduce apoi la observații noi, prezise și așa mai departe, către „adevărul" evaziv, la care, în general, nu ajungem niciodată. A întreba care este primul, teorie sau observație, este ca și cum ai întreba care vine primul, găina sau oul.

În general, progresele în cunoștințe în domeniul sănătății apar prin construirea, testarea și modificarea teoriilor. Epidemiologii fac inferențe inductive pentru a generaliza din multe observații , fac salturi creative ale imaginației pentru a deduce explicații și pentru a construi teorii și folosesc inferențe deductive pentru a testa acele teorii.

Prin urmare, teoriile pot fi folosite pentru a prezice observații. Dar

aceste observații nu vor fi întotdeauna exact așa cum le prezicem noi, din cauza erorii și a variabilității inerente a fenomenelor naturale. Dacă observațiile sunt foarte diferite de predicțiile noastre, va trebui să abandonăm sau să modificăm teoria. Cum testăm gradul de discordanță dintre predicțiile noastre bazate pe teorie și realitatea observațiilor noastre? Testul este un test statistic sau probabilistic. Este testul *ipotezei nule , care este piatra de temelie a inferenței statistice* și va fi discutată mai târziu. Câteva articole excelente despre logica și filozofia științei și aplicațiile în epidemiologie sunt enumerate în referințele de la sfârșitul acestei cărți. [2-6]

1.2 Variabilitatea fenomenelor necesită o analiză statistică

Statistica este o metodologie cu domenii largi de aplicare în știință și industrie, precum și în medicină și în multe alte domenii. Un fenomen se poate baza în principal pe un model determinist. Un exemplu este legea lui Boyle, care afirmă că, pentru un volum fix, o creștere a temperaturii unui gaz determină că există o creștere a presiunii . De fiecare dată când această lege este testată, apare același rezultat. Singura variabilitate constă în eroarea de măsurare. Multe fenomene din fizică și chimie sunt de o astfel de natură.

Un alt tip de model este un model probabilistic, care presupune că diverse stări ale unui fenomen apar cu anumite probabilități. De exemplu, distribuția inteligenței este în principal probabilistică, adică valorile date ale inteligenței apar cu o anumită probabilitate în populația generală. În biologie, psihologie sau medicină, unde fenomenele sunt influențate de mulți factori care în sine sunt variabili și de alți factori care nu sunt identificabili, modelele sunt adesea probabilistice . De fapt, pe măsură ce cunoștințele în fizică au devenit mai rafinate, începe să pară că modelele considerate anterior a fi deterministe sunt probabiliste.

În orice caz, atunci când modelul este în principal probabilistic, sunt necesare tehnici statistice pentru a crește cunoștințele științifice. *Prezența variației necesită utilizarea analizei statistice.* [1] Când există o variație mică în raport cu un fenomen, se acordă mult mai multă greutate unei cantități mici de dovezi decât atunci când există o mare variație. De exemplu, știm că cancerul pancreatic pare să fie în mod variabil o boală fatală. Astfel, dacă am găsi un medicament care a vindecat indiscutabil câțiva pacienți de

cancer pancreatic, am acorda multă greutate dovezilor că medicamentul ar reprezenta un leac, mult mai mult decât dacă evoluția acestei boli ar fi mai variabilă. Spre deosebire de acest exemplu , dacă am încerca să stabilim dacă vitamina C vindecă răceala, ar trebui să-i demonstrăm efectul la mulți pacienți și ar trebui să folosim metode statistice pentru a face acest lucru, deoarece ființele umane sunt destul de variabile în ceea ce privește răcelii. De fapt, în majoritatea fenomenelor biologice și cu atât mai mult în fenomenele sociale și psihologice, există o mare variabilitate.

1.3 Inferență inductivă: Statistica ca tehnologie a metodei științifice

Metodele statistice sunt metode obiective prin care *tendințele de grup sunt extrase din observațiile pe mai mulți indivizi separati.* Un concept simplu de statistică este calcularea mediilor, procentelor și așa mai departe și prezentarea datelor în tabele și diagrame. Astfel de tehnici de rezumare a datelor sunt într-adevăr foarte importante și esențiale pentru descrierea populației studiate. Cu toate acestea, ele constituie o mică parte din domeniul statisticii. O mare parte a statisticilor implică *extragerea de inferențe din eșantioane către o populație* cu privire la o caracteristică de interes. Să presupunem că suntem interesați de tensiunea arterială medie a femeilor studente. Dacă am putea măsura

 pentru fiecare membru al acestei populații, nu ar trebui să deducem nimic. Pur și simplu am face o medie a tuturor numerelor pe care le-am obținut. În practică, totuși, luăm un eșantion de studenți (corespunzător selectat), iar pe baza datelor pe care le obținem din eșantion, deducem care ar fi probabil media întregii populații.

Fiabilitatea unor astfel de inferențe sau concluzii poate fi evaluată în termeni de declarații de probabilitate. *În raționamentul statistic, deci, facem inferențe inductive, de la particular (eșantion) la general (populație).* Astfel, se poate spune că statistica este tehnologia metodei științifice.

1.4 Proiectarea Studiilor

În timp ce generarea de ipoteze poate proveni din observații anecdotice, testarea acelor ipoteze trebuie făcută prin observații controlate, fără părtiniri sistematice. Tehnicile statistice, pentru a fi valide, trebuie aplicate datelor obținute din studii bine concepute. În caz contrar, cunoștințele solide nu sunt avansate.

Există două tipuri de studii: (1) Studii observaționale, în care „Natura" determină cine este expus la factorul de interes și cine nu este expus. Aceste studii demonstrează asocierea. Asocierea poate implica cauzalitate sau poate nu. (2) Studii experimentale, în care investigatorul stabilește cine este expus. Acestea pot dovedi cauzalitate.

Studiile observaționale pot avea trei modele de studiu diferite: *transversale, caz-control* sau *prospective* . Într-un *studiu transversal,* măsurătorile sunt luate la un moment dat. De exemplu, într-un studiu transversal al hipertensiunii arteriale și al bolii coronariene, anchetatorii determină tensiunea arterială și prezența bolilor de inimă în același timp. Dacă găsesc o asociație, nu ar putea spune care a fost primul. Boala de inimă duce la hipertensiune arterială sau hipertensiunea arterială cauzează boli de inimă sau sunt atât hipertensiunea arterială, cât și bolile de inimă rezultatul unei alte cauze comune?

Într-un *studiu caz-control* al fumatului și al cancerului pulmonar, de exemplu, investigatorul începe cu cazurile de cancer pulmonar și cu controale, iar prin examinarea înregistrărilor sau prin interviuri determină prezența sau absența factorului de care este interesat (fumatul). Un studiu caz-control este uneori denumit un *studiu retrospectiv* deoarece datele despre factorul de interes sunt colectate retrospectiv și, prin urmare, pot fi supuse diferitelor inexactități.

Într-un studiu *prospectiv (* sau *de cohortă),* investigatorul începe cu o cohortă de persoane neboli cu acel factor (adică, cei care fumează) și persoane fără acel factor (nefumători), și continuă într-un timp viitor pentru a determina frecvența dezvoltării bolii în cele două grupuri. Un studiu prospectiv este cunoscut și sub denumirea de studiu longitudinal. *Distincția dintre studiile caz-control și studiile prospective constă în eșantionare. În studiul caz-control eșantionăm dintre cei bolnavi și nebolnavi, în timp ce într-un studiu prospectiv eșantionăm dintre cei cu factorul și cei fără factorul* . Studiile prospective oferă dovezi mai puternice de cauzalitate decât studiile retrospective, dar sunt adesea mai dificile, mai costisitoare și uneori imposibil de efectuat, de exemplu dacă boala studiată durează decenii pentru a se dezvolta sau dacă este foarte rară.

În domeniul sănătății, un studiu experimental pentru a testa o intervenție de un fel se numește *studiu clinic.* Într-un studiu clinic, investigatorul repartizează pacienții sau participanții într-un grup sau altul, de obicei aleatoriu, încercând să mențină toți ceilalți factori constanti sau controlați și compară rezultatul de interes în cele două (sau mai multe) grupuri. Mai multe despre studiile clinice găsiți în capitolul 6.

În rezumat, deci, următoarea listă este în ordinea crescătoare a puterii în ceea ce privește demonstrarea cauzalității:

- *studii transversale:* utile în arătarea asocierilor, în furnizarea de indicii timpurii asupra etiologiei.

- *studii caz-control:* utile pentru boli sau afecțiuni rare sau când boala durează foarte mult până se manifestă (denumire sinonimă: *studii retrospective*).

- *studii de cohortă:* utile pentru furnizarea de dovezi mai puternice de

cauzalitate și mai puțin supuse unor părtiniri din cauza erorilor de reamintire sau de măsurare (nume sinonime: *studii prospective, studii longitudinale).*

- *studii clinice:* studii prospective, experimentale, care oferă cele mai riguroase dovezi de cauzalitate.

1.5 Cum se cuantifică variabilele

Cum testăm o ipoteză? În primul rând, trebuie să stabilim ipoteza într-o manieră *cantitativă* . Măsura noastră de criteriu trebuie să fie un anumit număr. De exemplu, câți pacienți au murit într-un grup de medicamente în comparație cu câți dintre pacienții care nu au primit medicamentul au murit sau care este tensiunea arterială medie a pacienților tratați cu un anumit medicament antihipertensiv în comparație cu tensiunea arterială medie a pacienților care nu iau acest medicament. Uneori variabilele sunt greu de cuantificat. De exemplu, dacă evaluezi calitatea îngrijirii într-o clinică dintr-un spital în comparație cu clinica altui spital, poate fi uneori dificil să găsești o măsură cantitativă care să fie reprezentativă pentru calitatea îngrijirii, dar totuși se poate face și trebuie făcută dacă se dorește testarea ipotezei.

Există două tipuri de date cu care ne putem ocupa: *variabile discrete* sau *categorice* și *variabile continue.* Variabilele continue, teoretic, pot lua un număr infinit de valori între oricare două puncte fixe. De exemplu, greutatea este o variabilă continuă, la fel ca tensiunea arterială, timpul, inteligența și, în general, variabilele în care pot fi luate măsurători. Variabilele discrete (sau variabilele categorice) sunt variabile care pot lua doar anumite valori numerice fixe. De exemplu, sexul este o variabilă discretă. Îl puteți codifica ca 1 = bărbat, 2 = femeie bărbat, dar o persoană nu poate avea un cod de 1,5 pentru sex (cel puțin nu teoretic). Variabilele discrete se referă în general la numărare, cum ar fi numărul de pacienți dintr-un anumit grup care trăiesc, numărul de persoane cu o anumită boală și așa mai departe. În capitolul 3 vom lua în considerare o tehnică de testare a unei ipoteze în care variabila este una discretă și

ulterior, vom discuta unele aspecte ale variabilelor continue, dar mai întâi vom discuta conceptele generale ale testării ipotezelor.

1.6 Ipoteza nulă

Ipoteza pe care o testăm statistic se numește *ipoteza nulă.* Să luăm un exemplu conceptual simplu. Să presupunem că testăm eficacitatea unui nou medicament la pacienții cu infarct miocardic (atac de cord). Împărțim pacienții în două grupuri - medicament și fără medicament - conform procedurilor de proiectare bune și folosim ca criteriu de măsurare a mortalității în cele două grupuri. Sperăm că medicamentul scade mortalitatea, dar pentru a testa ipoteza statistic, trebuie să o punem într-un fel invers. Spunem că ipoteza noastră este că medicamentul nu face nicio diferență și ceea ce sperăm să facem este să respingem ipoteza „nicio diferență", pe baza dovezilor din eșantionul nostru de pacienți. Aceasta este cunoscută sub numele de *ipoteza nulă* . Specificăm ipoteza noastră de testare după cum urmează:

H_O (ipoteza nulă): rata mortalității în grupul tratat cu medicamentul A = rata mortalității în grupul tratat cu medicamentul B.

Aceasta este echivalentă cu:

H_O : (rata mortalității în grupa A) - (rata mortalității în grupa B) = 0.

Testăm acest lucru față de o *ipoteză alternativă,* cunoscută sub numele de H_A , conform căreia diferența în ratele de deces între cele două grupuri *nu este* egală cu 0.

Apoi colectăm date și notăm diferența *observată* în mortalitate între grupul A și grupul B. Dacă această diferență observată este suficient de mai mare decât zero, respingem ipoteza nulă. Dacă respingem ipoteza nulă a lipsei de diferență, acceptăm *ipoteza alternativă* , care este că medicamentul face o diferență.

Când testați o ipoteză, acesta este tipul de raționament pe care îl utilizați:

(1) Voi *presupune că* ipoteza că nu există nicio diferență este adevărată;

(2) Voi colecta apoi datele și *voi observa* diferența dintre cele două grupuri;
(3) Dacă ipoteza nulă este adevărată, cât de probabil este ca *doar întâmplător* să obțin rezultate ca acestea?
(4) Dacă nu este probabil ca aceste rezultate să apară întâmplător în ipoteza că ipoteza nulă este adevărată, atunci voi concluziona că este falsă și voi „accepta" ipoteza alternativă.

1.7 De ce testăm ipoteza nulă?

Să presupunem că credem că medicamentul A este mai bun decât medicamentul B în prevenirea decesului în urma unui atac de cord. De ce nu testăm acea credință în mod direct și să vedem care medicament este mai bun, în loc să testăm ipoteza că medicamentul A este *egal* cu medicamentul B? Motivul este că există un număr infinit de moduri în care medicamentul A poate fi mai bun decât medicamentul B, așa că ar trebui să testăm un număr infinit de ipoteze. Dacă medicamentul A provoacă cu 10% mai puține decese decât medicamentul B, este mai bine. Deci mai întâi ar trebui să vedem dacă medicamentul A provoacă cu 10% mai puține decese. Dacă nu provoacă decese cu 10% mai puține, dar dacă provoacă decese cu 9% mai puține, este și mai bine. Apoi ar trebui să testăm dacă observațiile noastre sunt în concordanță cu o diferență de 9% a mortalității între cele două medicamente. Apoi ar trebui să testăm dacă există o diferență de 8% și așa mai departe. Notă: fiecare astfel de ipoteză ar fi creată ca o ipoteză nulă în următoarea formă: Medicamentul A - Mortalitatea medicamentului B = 10% sau echivalent,

(Medicament A - Mortalitatea medicamentului B) - (10%) = 0;
(Medicament A - Mortalitatea medicamentului B) - (9%) = 0;
(Medicament A - Mortalitatea medicamentului B) - (8%) = 0;
etc.

Pe de altă parte, atunci când testăm ipoteza nulă a nediferențelor , trebuie să testăm doar o singură valoare – o diferență de 0% – și ne întrebăm dacă observațiile noastre sunt în concordanță cu ipoteza că *nu există nicio* diferență de mortalitate între cele două medicamente. Dacă observațiile sunt

în concordanță cu o diferență nulă, atunci nu putem afirma că un medicament este mai bun decât celălalt. Dacă este puțin probabil ca acestea să fie în concordanță cu o diferență nulă, atunci putem respinge acea ipoteză și putem concluziona că există o diferență.

O sursă comună de confuzie apare atunci când investigatorul dorește cu adevărat să arate că un tratament este la fel de bun ca altul (spre deosebire de exemplul de mai sus, în care investigatorul din inima ei crede cu adevărat că un medicament este mai bun). De exemplu, în camera de urgență s-ar putea să fi fost concepută o procedură mai rapidă și anchetatorul consideră că poate fi la fel de bună ca procedura standard, care durează mult timp. Tentația într-o astfel de situație este „demonstrarea ipotezei nule". *Dar este imposibil să „demonstrezi" ipoteza nulă.*

Toate testele statistice pe care le pot face este să respingă ipoteza nulă sau să nu o respingă. Nu dovedim ipoteza prin adunarea de dovezi afirmative sau de susținere, deoarece indiferent de câte ori am făcut experimentul și am găsit o diferență aproape de zero, nu am putea fi niciodată siguri că data viitoare când vom face un astfel de experiment nu vom găsi o diferență uriașă care să nu se apropie de zero. Este ca exemplul lebedelor albe discutat mai devreme: indiferent câte lebede albe vedem, nu putem dovedi că toate lebedele sunt albe, deoarece următoarea observare ar putea fi o lebădă neagră. Mai degrabă, încercăm să falsificăm sau să respingem afirmația noastră de lipsă de diferență și, dacă afirmația de zero diferență rezistă tentativei noastre de respingere, ea supraviețuiește ca o ipoteză în care continuăm să avem credință. Nerespectarea acesteia nu înseamnă că am demonstrat că nu există într-adevăr nicio diferență. Înseamnă pur și simplu că dovezile pe care le avem „sunt în concordanță cu" ipoteza nulă. Rezultatele pe care le-am obținut ar fi putut apărea doar întâmplător dacă ipoteza nulă ar fi fost adevărată. (Poate că designul studiului nostru nu a fost adecvat. Poate că nu am avut destui pacienți.)

Deci, ce se poate face dacă cineva dorește cu adevărat să arate că două tratamente sunt echivalente? *Se poate proiecta un studiu care este suficient de mare pentru a detecta o mică diferență dacă există într-adevăr una.* Dacă studiul are puterea (adică o probabilitate mare) de a detecta o diferență care este foarte, foarte, foarte mică și nu reușește să o detecteze, atunci se poate spune cu un grad ridicat de încredere că nu se poate găsi o diferență semnificativă între cele două tratamente. Este imposibil să existe un studiu

cu putere suficientă pentru a detecta o diferență de 0%. Pe măsură ce diferența pe care se dorește să se detecteze se apropie de zero, numărul de subiecți necesar pentru o anumită putere se apropie

infinit. Relațiile dintre nivelul de semnificație, putere și dimensiunea eșantionului sunt discutate mai pe larg în Capitolul 6.

1.8 Tipuri de erori

Ideea importantă este că *nu putem fi niciodată siguri* că avem dreptate fie când acceptăm, fie că respingem o ipoteză. De fapt, riscăm să facem unul dintre cele două tipuri de erori: putem respinge ipoteza nulă sau testarea incorect, adică putem concluziona că medicamentul reduce mortalitatea atunci când de fapt nu are niciun efect. Aceasta este cunoscută ca eroare de tip *I.* Sau putem să nu respingem incorect ipoteza nulă sau de testare, adică putem concluziona că medicamentul nu are efect când de fapt reduce mortalitatea. Aceasta este cunoscută sub numele de *eroare de tip II.* Fiecare dintre aceste erori poartă cu sine anumite consecințe. În unele cazuri, o eroare de tip I poate fi mai gravă; în alte cazuri o eroare de tip II poate fi mai gravă. Aceste puncte sunt ilustrate în Figura 1.2.

Ipoteza nulă ($HO_)$: *Drogul nu are efect* - nu există diferențe de mortalitate între pacienții care utilizează medicamente și pacienții care nu le folosesc.

Ipoteza alternativă ($HA_)$: *Medicamentul are efect* - reduce mortalitatea.

		ADEVĂRATĂ STARE A NATURII	
		DROGUL NU ARE EFECT H_O Adevărat	DROGUL ARE EFECT; H_O Fals, H_E adevărat
DECIZIE PE BAZĂ DE PROBA	NU RESPINGE H_O Niciun efect NU	NU EROARE	TIP II EROARE
	RESPING H_O (Accept H_A) Efect	TIP I EROARE	NU EROARE

Figura 1.2

Dacă nu respingem H_O, ajungem la concluzia că nu există nicio relație între drog și mortalitate. Dacă respingem H_O și acceptăm HA, ajungem la concluzia că există o relație între drog și mortalitate.

Acțiuni care trebuie întreprinse pe baza deciziei:

(1) Dacă credem ipoteza nulă (adică, nu reușim să o respingem), nu vom folosi medicamentul.
*Consecințele unei decizii **greșite** :* eroare de tip II. Dacă credem incorect H_O, deoarece în realitate medicamentul este benefic, reținându-l vom permite să moară pacienții care altfel ar fi putut trăi.

(2) Dacă respingem ipoteza nulă în favoarea ipotezei alternative, vom folosi medicamentul.
*Consecințele unei decizii **greșite** :* eroare de tip I. Dacă am respins incorect ipoteza nulă, vom folosi medicamentul și pacienții nu beneficiază. Presupunând că medicamentul nu este dăunător în sine, nu rănim direct pacienții, dar din moment ce credem că am găsit leacul, s-ar putea să nu mai testăm alte medicamente.

Nu putem niciodată să cunoaștem în mod absolut „Starea adevărată a naturii", dar o deducem pe baza unor probe de dovezi .

1.9 Nivel de semnificație și tipuri de eroare

Nu putem elimina riscul de a face unul dintre aceste tipuri de erori, dar putem reduce probabilitățile de a face aceste erori. *Probabilitatea de a face o eroare de tip I este cunoscută ca nivelul de semnificație al unui test statistic.* Când citiți în literatură că un rezultat a fost semnificativ la nivelul .05 înseamnă că în acest experiment rezultatele sunt astfel încât probabilitatea de a face o eroare de tip I este mai mică sau egală cu .05. În cea mai mare parte, în experimente și sondaje, oamenii sunt foarte - îngrijorați că au o probabilitate scăzută de a face o eroare de tip I și din zece ignoră eroarea de tip II. Aceasta poate fi o greșeală, deoarece în unele cazuri o eroare de tip II este una mai gravă decât o eroare de tip I. În proiectarea unui studiu, dacă doriți să reduceți eroarea de tip I, creșteți automat valoarea

 probabilitatea de eroare de tip II. Pentru a reduce probabilitățile erorii de tip I și de tip II într-un studiu, este necesar să creștem numărul de observații.

Este interesant de observat că regulile Administrației pentru Alimente și Medicamente (FDA) sunt stabilite pentru a reduce probabilitatea de a face erori de tip I. Pentru ca un medicament să fie aprobat pentru comercializare, compania de medicamente trebuie să fie capabilă să demonstreze că nu dăunează și că este eficient. Astfel, multe medicamente sunt respinse deoarece eficacitatea lor nu poate fi demonstrată în mod adecvat. Ipoteza nulă testată este: „Acest medicament nu are nicio diferență". Pentru a satisface regulile FDA, această ipoteză trebuie respinsă, probabilitatea de a face o eroare de tip I (adică de a o respinge incorect) fiind destul de scăzută. Cu alte cuvinte, FDA nu vrea multe medicamente inutile pe piață. Companiile de medicamente, cu toate acestea, acordă importanță și protecției împotriva erorii de tip II (adică să evite să creadă incorect ipoteza fără diferențe), astfel încât să poată comercializa medicamente potențial benefice.

1.10 Consecințele erorilor de tip I și tip II

Severitatea relativă a acestor erori depinde de situație. Rețineți că o eroare de tip I (cunoscută și sub numele de *alfa*) înseamnă că spuneți că ceva este într-adevăr acolo (un efect) atunci când de fapt nu este, iar o eroare de tip II (cunoscută și ca eroare *beta*) înseamnă că vă lipsește ceva care este cu adevărat acolo. Dacă cauți un remediu pentru cancer, o eroare de tip II ar fi destul de gravă. V-ar lipsi să găsiți tratamente utile . Dacă vă gândiți la un medicament scump pentru a trata o răceală, în mod clar ați dori să evitați o eroare de tip I, adică nu ați dori să faceți afirmații false pentru un remediu pentru răceală.

Este greu de reținut distincția dintre erorile de tip I și II. Poate că această mică pildă ne va ajuta. A fost odată un rege care era foarte gelos pe regina lui. Avea doi cavaleri, Alpha, care era foarte frumos, și Beta, care era foarte urâtă. S-a întâmplat ca Regina să fie îndrăgostită de Beta. Regele, însă, a bănuit că regina avea o aventură cu Alpha și l-a decapitat. Astfel, Regele a făcut ambele tipuri de erori: a bănuit o relație (cu

Alpha) unde nu era niciunul și nu a reușit să detecteze o relație (cu Beta) în care chiar exista una. Regina a fugit din regat împreună cu Beta și a trăit fericiți pentru totdeauna, în timp ce Regele a suferit chinuri de vinovăție din cauza respingerii greșite și fatale a lui Alpha.

Mai multe despre alfa, beta, putere și dimensiunea eșantionului apar în Capitolul 6. Deoarece testarea ipotezelor se bazează pe probabilități, vom prezenta mai întâi câteva concepte de bază de probabilitate în Capitolul 2.

Capitolul 2
PUTIN PROBABILITATE

Teoria probabilității nu este, în fond, altceva decât bunul simț redus la calcul.
Pierre Simon De Le Place Theori Analytique des Probabilites (1812-1820)

2.1 Ce este probabilitatea?

Probabilitatea de apariție a unui eveniment este indicată printr-un număr cuprins între 0 și 1. Un eveniment a cărui probabilitate de apariție este 0 este sigur că nu va avea loc, în timp ce un eveniment a cărui probabilitate este 1 este sigur că va avea loc.

Definiția clasică a probabilității este următoarea: dacă un eveniment poate avea loc în *N* moduri care se exclud reciproc, la fel de probabile și dacă n_A dintre aceste rezultate au atributul *A,* atunci probabilitatea lui *A* , scrisă ca *P(A),* este egală cu n_A / N . Aceasta este o definiție a priori a probabilității, adică se determină probabilitatea unui eveniment înainte ca acesta să se întâmple. Să presupunem că cineva ar trebui să arunce un zar și să știe probabilitatea de a obține un număr divizibil cu trei la aruncarea unui zar. Există șase moduri posibile prin care zarul poate ateriza. Dintre acestea, există două moduri în care numărul de pe fața zarului este divizibil cu trei, un 3 și un 6. Astfel, probabilitatea de a obține un număr divizibil cu trei la aruncarea unui zar este 2/6 sau 1/3.

În multe cazuri, totuși, nu suntem capabili să enumerăm toate modurile posibile în care poate apărea un eveniment și, prin urmare, folosim definiția frecvenței *relative a probabilității.* Acesta este definit ca numărul de ori în care evenimentul de interes a avut loc împărțit la numărul total de încercări (sau oportunități pentru ca evenimentul să aibă loc). Deoarece se bazează pe date anterioare, se numește *definiția a posteriori a probabilității.*

De exemplu, dacă selectați la întâmplare o femeie americană albă, probabilitatea ca ea să moară de boli de inimă este de 0,00287. Aceasta se bazează pe constatarea că la 100.000 de femei albe americane, 287 au murit din cauza bolii coronariene (estimările sunt pentru 2001, Centrul Național de Statistică în Sănătate 7) . Când luați în considerare probabilitatea unei

femei americane albe cu vârsta cuprinsă între 45 și 64 de ani, cifra scade la 0,00088 (sau 88 de femei din grupa respectivă de vârstă din 100.000), iar când luați în considerare femeile de 65 de ani și mai mult, cifra crește la 0,01672 (sau 1672 la 100.000). Pentru bărbații albi de 65 de ani sau mai mult, este .0919 (sau 9190 la 100.000). Cele două puncte importante sunt (1) pentru a determina o probabilitate, *trebuie să specificați populația la care vă referiți,* de exemplu, toți masculii albi femele , masculii albi între 65 și 74 de ani, femelele nealbe între 65 și 74 de ani și așa mai departe; și (2) *cifrele de probabilitate sunt revizuite în mod constant* pe măsură ce devin disponibile noi date.

Acest lucru ne aduce la noțiunea de *frecvență așteptată* . Dacă probabilitatea unui eveniment este *P* și există *N* încercări (sau oportunități ca evenimentul să se producă), atunci ne putem aștepta ca evenimentul *să* se producă *de N* x *P* ori. Este necesar să ne amintim că probabilitatea „funcționează" pentru numere mari. Când aruncăm o monedă spunem că probabilitatea ca aceasta să cadă în cap este de .50, ne referim la faptul că în multe aruncări, jumătate din timp moneda va ateriza capete. Dacă aruncăm moneda de zece ori, putem obține trei capete (30%) sau șase capete (60%), ceea ce reprezintă o abatere considerabilă de la 50% la care ne așteptăm. Dar dacă aruncăm moneda de 200.000 de ori, este foarte probabil să fim aproape de a obține exact 100.000 de capete sau 50%.

Frecvența așteptată este într-adevăr modul în care probabilitatea „funcționează". Este dificil de conceptualizat aplicarea probabilității unui individ. De exemplu, când cranicii TV proclamă că vor fi, de exemplu, 400 de accidente mortale în statul X pe 4 iulie, este imposibil să spunem dacă o persoană individuală va avea într-adevăr un astfel de accident, dar putem fi destul de siguri că numărul acestor accidente va fi foarte aproape de cele 400 prezise (pe baza probabilităților derivate din statisticile anterioare din 4 iulie).

2.2 Combinarea probabilităților

Există două legi pentru combinarea probabilităților care sunt importante. În primul rând, dacă există *evenimente care se exclud reciproc* (adică, dacă unul are loc, celălalt nu poate), probabilitatea ca unul sau celălalt să se producă este *suma* probabilităților lor individuale. Simbolic,

$$P(A \text{ sau} B) = P(A) + P(B)$$

Un exemplu în acest sens este următorul: probabilitatea de a obține fie un 3, fie un 4 la aruncarea unui zar este 1/6 + 1/6 = 2/6.

Un lucru util de știut este că suma probabilităților individuale ale tuturor evenimentelor posibile care se exclud reciproc trebuie să fie egală cu 1. De exemplu, dacă *A* este evenimentul câștigării la loterie, și nu *A* (scris ca *A*), este evenimentul necâștigării la loterie, atunci *P* (*A*) + *P* (*A*) = 1,0 și *P* (A (*P*) = 1 - *A* (*P*) =).

În al doilea rând, dacă există două evenimente independente (adică, apariția unuia nu este legată de apariția celuilalt), probabilitatea comună ca acestea să se producă împreună (în comun) este *produsul* probabilităților individuale . Simbolic,

$$P(A \text{ și } B) = P(A) \times P(B)$$

Un exemplu în acest sens este probabilitatea ca la aruncarea unui zar să obțineți un număr care este atât par, cât și divizibil cu 3. Această probabilitate este egală cu 1/2 x 1/3 = 1/6. (Singurul număr par și divizibil cu 3 este numărul 6.)

Legea probabilității comune este utilizată pentru a testa dacă evenimentele sunt independente . Dacă sunt independenți, produsul probabilităților lor individuale ar trebui să fie egal cu probabilitatea comună. Dacă nu, ei nu sunt independenți . Este baza testului de semnificație chi-pătrat, pe care îl vom lua în considerare în secțiunea următoare.

Să aplicăm aceste concepte la un exemplu medical. Rata mortalității pentru cei cu infarct într-o unitate specială de îngrijire coronariană dintr-un anumit spital este de 15%. Astfel, probabilitatea ca un pacient cu infarct internat în această unitate coronariană să moară este de 0,15 și să supraviețuiască este de 0,85. Dacă doi bărbați sunt internați în unitatea de îngrijire coronariană într-o anumită zi, fie *A* evenimentul în care primul bărbat moare și fie *B* evenimentul în care al doilea bărbat moare.

Probabilitatea ca ambii să moară este

$$P(A \text{ și } B) = P(A) \times P(B) = .15 \times .15 = .0225$$

Presupunem că aceste evenimente sunt independente unele de altele, astfel încât să le putem multiplica probabilitățile. Rețineți, totuși, că probabilitatea ca unul *sau* altul să moară din cauza atacului de cord *nu este* suma probabilităților lor, deoarece aceste două evenimente nu se exclud reciproc. Este posibil ca ambii să moară (adică pot apărea atât *A* cât și *B*).

Pentru a face acest lucru mai clar, o modalitate bună de abordare a probabilității este prin utilizarea diagramelor Venn, așa cum se arată în Figura 2.1. Diagramele Venn constau din pătrate care reprezintă universul posibilităților și cercuri care definesc evenimentele de interes.

În diagramele 1, 2 și 3, spațiul din interiorul pătratului reprezintă toate *cele N* rezultate posibile. Cercul marcat cu A reprezintă toate rezultatele care constituie evenimentul *A* ; cercul marcat cu B reprezintă toate rezultatele care constituie evenimentul *B* . Diagrama 1 ilustrează două evenimente care se exclud reciproc; un rezultat din cercul A nu poate fi și în cercul B. Diagrama 2 ilustrează două evenimente care pot avea loc împreună: un rezultat din cercul A poate fi și un rezultat aparținând cercului B. Zona umbrită marcată AB reprezintă rezultate care sunt apariția ambelor *A și B.* Diagrama 3 reprezintă două evenimente în care unul (*B*) este un subset al celuilalt (*A*); un rezultat din cercul B trebuie să fie, de asemenea, un rezultat care constituie evenimentul *A* , dar inversul nu este neapărat adevărat.

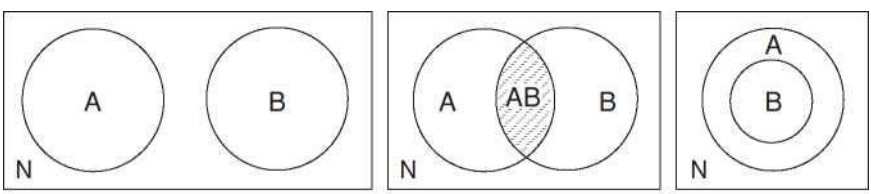


Figure 2.1

Din diagrama 2 se poate observa că dacă dorim ca probabilitatea ca un rezultat să fie *A sau B* și dacă adăugăm rezultatele din cercul A la rezultatele din cercul B, am adăugat rezultatele din zona umbrită de două ori. Prin

urmare, trebuie să scădem rezultatele din zona umbrită *(A și B)* scrisă și ca *(AB*) o dată pentru a ajunge la răspunsul corect . Astfel, obținem rezultatul

$$P(A \text{ sau } B) = P(A) + P(B) - P(AB)$$

2.3 Probabilitate condiționată

Acum să luăm în considerare cazul în care șansa ca un anumit eveniment să se întâmple depinde de rezultatul unui alt eveniment. Probabilitatea lui *A* , dat fiind că *B* a avut loc, se numește probabilitate condiționată a lui *A* dat *B* și este scrisă simbolic ca *P* (*A* | *B*). O ilustrare a acestui lucru este oferită de diagrama Venn 2. Când vorbim de probabilitatea condiționată, numitorul devine toate rezultatele din cercul B (în loc de toate *N* rezultate posibile), iar numărătorul este format din acele rezultate care se află în acea parte a *lui A* care conține și rezultatele aparținând lui *B* . Aceasta este zona umbrită din diagrama etichetată AB. Dacă ne întoarcem la definiția noastră originală a probabilității, vedem asta

$$P(A\backslash B) = \underset{n_B}{}$$

(numărul de rezultate atât în *A cât și în B* , împărțit la numărul total de rezultate în *B*).

Dacă împărțim atât numărătorul cât și numitorul la *N* , numărul total al *tuturor* rezultatelor posibile, obținem

$$P(A\backslash B) = \qquad \begin{matrix} n_{AB}/N \_ P(A \text{ și } B) \\ n_B/N \sim \quad P(B) \end{matrix}$$

Înmulțirea ambelor părți cu *P* (*B*) dă legea multiplicativă *completă* :

$$P(AandB) = P(A\backslash B) \times P(B)$$

Desigur, dacă *A* și *B* sunt independenți, atunci probabilitatea lui *A* dat *B* este doar egală cu probabilitatea lui *A* (deoarece apariția lui *B* nu influențează apariția lui *A*) și apoi vedem că

$$P(AandB) = P(A) \times P(B)$$

2.4 Probabilitatea Bayesiană

Imaginați-vă că *M* este evenimentul „pierderea memoriei", iar *B* este evenimentul „tumoare pe creier". Din cercetările efectuate asupra pacienților cu tumori cerebrale putem stabili probabilitatea *pierderii memoriei în cazul unei tumori pe creier, P(M\B).* Un clinician, însă, este mai interesat de probabilitatea *unei tumori cerebrale , având în vedere că un pacient are pierderi de memorie, P(B | M*).

Este dificil să se obțină acea probabilitate în mod direct, deoarece ar trebui să se studieze numărul mare de persoane cu pierderi de memorie (care în majoritatea cazurilor provine din alte cauze) și să se determine ce proporție dintre ele au tumori cerebrale.

Ecuația lui Bayes (sau teorema lui Bayes) estimează *P* (*B* | *M*) după cum urmează:

$$P(\text{tumoare cerebrală, dată fiind pierderea memoriei}) = \frac{P(\text{pierderea memoriei, dată fiind tumora cerebrală}) \times P(\text{tumoare cerebrală})}{P(\text{pierdere de memorie})}$$

La numitor, evenimentul „pierderii memoriei" poate apărea fie în rândul persoanelor cu tumoră cerebrală, cu probabilitate = *P* (*M* | *B*) *P* (*B*), fie în rândul persoanelor fără tumoră cerebrală, cu probabilitate = *P* (*M* | *B*) *P* (*B*). Astfel,

$$P(B\backslash M) = \frac{P(M\backslash B)P(B)}{P(M \backslash B)P(B) + P(M \backslash B)P(B)}$$

Probabilitatea generală a unei tumori cerebrale, *P* (*B*) este „probabilitatea a priori", care este un fel de „cea mai bună presupunere" a prevalenței tumorilor cerebrale.

2.5 Cote și probabilitate

Când șansele ca un anumit cal *să piardă* o cursă sunt de 4 la 1, el are o probabilitate de 4/5 = 0,80 de a pierde. Pentru a converti o declarație de cote în probabilitate, adăugăm 4 + 1 pentru a obține numitorul nostru de 5. Cotele ca calul *să câștige* sunt 1 la 4, ceea ce înseamnă că are o probabilitate de câștig de 1/5 = 0,20.

$$\textit{The odds in favor of } A = \frac{P(A)}{P(notA)} \quad \frac{P(A)}{1 - P(A)}$$

$$P(A) = \frac{odds}{1 + odds}$$

Șansele de a extrage un as = 4 (și într-un pachet) la 48 (cărți care nu suntași) = 1 la 12; prin urmare, *P* (as) = 1/13. Cota *împotriva* tragerii unui as = 12 la 1; *P* (Nu As) = 12/13.

În medicină, cotele sunt adesea folosite pentru a calcula un *raport de cote*. Un raport de cote este pur și simplu raportul dintre două cote. De exemplu, să spunem că într-un anumit studiu care a comparat pacienții cu cancer pulmonar cu martori, s-a descoperit că șansele de a fi un caz de cancer pulmonar pentru persoanele care fumează au fost de 5 până la 4 (5/4) și șansele de a avea cancer pulmonar pentru nefumători sunt de 1 la 8 (1/8), atunci cota de șanse ar fi

$$5/4 \_ \underline{5 \times 8} \_ \underline{40}$$
$$1/8 - 4 \times 1 \text{ " } 4$$

Un raport de șanse de 10 înseamnă că șansele de a fi un caz de cancer pulmonar este de 10 ori mai mare pentru fumători decât pentru nefumători.

Rețineți, totuși, că nu putem determina dintr-o astfel de analiză care este probabilitatea de a face cancer pulmonar pentru fumători, pentru că pentru a face asta ar trebui să știm câți oameni din toți fumătorii au dezvoltat cancer pulmonar și nu am studiat toți fumătorii; tot ce știm este câți dintre toate cazurile noastre de cancer pulmonar au fost fumători . Nici nu putem obține probabilitatea de cancer pulmonar în rândul nefumătorilor, pentru că ar trebui să ne uităm la o populație de nefumători și să vedem câți dintre ei

au dezvoltat cancer pulmonar. Tot ce știm este că fumătorii au șanse de 10 ori mai mari de a avea cancer pulmonar decât nefumătorii .

Mai multe despre acest subiect sunt prezentate în Secțiunea 4.12.

2.6 Raportul de probabilitate

Un concept înrudit este raportul de probabilitate (LR), care ne spune cât de probabil este ca un anumit rezultat să apară dintr-un set de circumstanțe în raport cu cât de probabil ar apărea rezultatul dintr-un set opus de circumstanțe .

De exemplu, dacă un pacient are o pierdere bruscă a memoriei, am putea dori să cunoaștem probabilitatea acelui simptom pentru o tumoare pe creier, de exemplu. Ceea ce vrem să știm este probabilitatea ca pierderea de memorie să apară din tumora cerebrală *în raport cu* probabilitatea ca aceasta să apară dintr-o altă afecțiune. Raportul de probabilitate este un raport al probabilităților condiționate.

$$LR = \frac{P(\text{pierderea memoriei, dată fiind tumora cerebrală})}{P(\text{pierderea memoriei, fără tumori cerebrale})}$$

$$\frac{P(M \text{ dat } B)}{P(M \text{ dat nu } B)}$$

Desigur, pentru a calcula acest LR ar trebui să avem estimări ale probabilităților implicate în ecuație, adică ar trebui să știm următoarele: în rândul persoanelor care au tumori cerebrale, ce proporție au pierderi de memorie și între persoanele care nu au tumori cerebrale, ce proporție au pierderi de memorie. Uneori poate fi destul de dificil să stabilim numitorul raportului de probabilitate, deoarece ar trebui să cunoaștem prevalența pierderii memoriei în populația generală.

LR este poate mai practic de utilizat decât teorema lui Bayes, care oferă probabilitatea unei anumite boli având în vedere un anumit

simptom. În orice caz, este utilizat pe scară largă într-o varietate de situații, deoarece abordează această întrebare importantă: dacă un pacient prezintă un simptom, care este probabilitatea ca simptomul să se datoreze unei anumite boli, mai *degrabă decât* unui alt motiv decât această boală?

2.7 Rezumatul probabilității

Legea aditivă:

$$P(A \text{ sau } B) = P(A) + P(B) - P(A \text{ și } B)$$

Dacă evenimentele se exclud reciproc: $P(A \text{ sau } B) = P(A) + P(B)$.

Legea multiplicativă:

$$P(A \text{ și } B) = P(A \backslash B) \times P(B)$$

Dacă evenimentele sunt independente: $P(AandB) = P(A) \times P(B)$.

Probabilitate condiționată:

$$P(A \backslash B) = \frac{P(A \text{ and } B}{)}$$

Bayes' Theorem:

$$P(B \backslash A) = \frac{P(A \backslash B)P(B)}{P(A \backslash B)\,P(B) + P(A \backslash B)\,P(B)}$$

Odds of A:

$$\frac{P(A)}{\mathrm{P}(\mathrm{A})}$$

Raportul de probabilitate:

$$P(A\backslash B)$$
$$P(A\backslash B)$$

Capitolul 3
MAI MULT DESPRE STATISTICĂ

Un statistician este cineva care, cu capul într-un cuptor și picioarele într-o găleată cu apă cu gheață, la întrebarea cum se simte, răspunde: „În medie, mă simt bine”.

Anonim

Diferite tehnici statistice sunt adecvate în funcție de faptul că variabilele de interes sunt discrete sau continue. Vom lua în considerare mai întâi cazul variabilelor discrete și vom prezenta testul chi-pătrat și apoi vom discuta metode aplicabile variabilelor continue.

3.1 Chi-Pătrat pentru 2 x 2 Mese

Testul chi-pătrat este o metodă statistică pentru a determina dacă rezultatele unui experiment pot apărea întâmplător sau nu. Să luăm, prin urmare, exemplul testării unui medicament anticoagulant pe paciente de sex feminin cu infarct miocaral. Sperăm că medicamentul scade mortalitatea, dar am stabilit ipoteza noastră nulă după cum urmează:

- Nulă Ipoteză — Nu există nicio diferență în ceea ce privește mortalitatea între grupul de pacienți tratat și grupul de control.
- Alternativ Ipoteză: — Mortalitatea în lotul tratat este mai mică decât în grupul martor.

(Datele pentru exemplul nostru provin dintr-un studiu realizat cu mult timp în urmă și se referă la un grup specific cu risc ridicat. [8] Sunt utilizate în scopuri ilustrative și nu reflectă ratele actuale de mortalitate pentru persoanele cu infarct miocardic.)

Apoi ne înregistrăm datele într-un tabel *de urgență 2* x 2 în care fiecare pacient este clasificat ca aparținând uneia dintre cele patru celule:

	Frecvențele observate		
	Controla	Tratat	
A trăit	89	223	312
A murit	40	39	79
Total	129	262	391

Mortalitatea la lotul martor este de 40/129 = 31% iar la tratat este de 39/262 = 15%. Dar ar fi putut această diferență să fi apărut întâmplător? Folosim testul chi-pătrat pentru a răspunde la această întrebare. Ceea ce ne întrebăm cu adevărat este dacă cele două categorii de clasificare (control vs. tratat prin trăit vs. murit) sunt independente una de cealaltă. Dacă sunt independente, la ce frecvențe ne-am aștepta în fiecare dintre celule? Și *cât de diferite sunt frecvențele noastre observate de cele așteptate?* Cum măsurăm dimensiunea diferenței?

Pentru a determina frecvențele așteptate, luați în considerare următoarele:

	Controla	Tratat	
A trăit	*o*	*b*	*(a + b)*
A murit	*c*	*d*	*(c + d)*
Total	*(a + c)*	*(b + d)*	*N*

Dacă categoriile sunt independente, atunci probabilitatea ca un pacient să fie atât control, cât și viu este *P* (control) x *P* (trăit). [Aici aplicăm legea menționată în capitolul 2 privind probabilitatea comună a două evenimente independente.]

Frecvența așteptată a unui eveniment este egală cu probabilitatea evenimentului înmulțit cu numărul de încercări = *N* x *P* . Deci numărul așteptat de pacienți care sunt *atât* martori *, cât și* vii este

$$N \times P(\text{control și trăit}) = N \times P(\text{control}) \times P(\text{trăit})$$

$$\frac{(aa+A \times (aa+\wedge l}{NN} = \frac{{}_{(}a + {}_{c)} \times (aa+\wedge l}{N}$$

În cazul nostru, rezultă următorul tabel:

	Control	Treated	
Lived	$129 \times \frac{312}{391} = 103$	$262 \times \frac{312}{391} = 209$	312
Died	$129 \times \frac{79}{391} = 26$	$262 \times \frac{79}{391} = 53$	79
Total	129	262	391

Un alt mod de a privi acest lucru este de a spune că, deoarece 80% dintre pacienții din totalul studiului au trăit (adică, 312/391 = 80%), ne-am aștepta ca 80% dintre pacienții de control și 80% dintre pacienții tratați să trăiască. Aceste așteptări diferă, după cum vedem, de frecvențele observate notate mai devreme, adică acei pacienți tratați au avut, de fapt, o mortalitate mai mică decât cei din grupul de control.

Ei bine, acum că avem un tabel cu frecvențe observate și un tabel cu valori așteptate, de unde știm cât de diferite sunt acestea? Diferă ele doar întâmplător sau există vreun alt factor care le face să difere? Pentru a determina acest lucru, calculăm o valoare numită chi-pătrat (scrisă și ca Z^2). Aceasta se obține prin luarea valorii observate în fiecare celulă, scăzând din aceasta valoarea așteptată în fiecare celulă, pătrarea acestei diferențe și împărțirea la valoarea așteptată pentru fiecare celulă. Când se face acest lucru pentru fiecare celulă, cele patru cantități rezultate sunt adăugate împreună pentru a da un număr numit chi-pătrat. Din punct de vedere simbolic, această formulă este următoarea:

$$\frac{(O_a - e_a)^2}{£_a} + \frac{(O_b \sim e_b)^2}{e_b} + \frac{(O_c \sim e_c)^2}{e_c} + \frac{(O_d - e_d)^2}{e_d}$$

unde *O* este frecvența observată și *e* este frecvența așteptată în fiecare celulă.

Acest număr, numit chi-pătrat, este o statistică care are o distribuție cunoscută. Ceea ce înseamnă, în esență, este că pentru un număr infinit de astfel de tabele 2 x 2, s-au calculat chi-pătrat și astfel știm care este

probabilitatea de a obține anumite valori ale chi-pătratului. Astfel, atunci când calculăm un chi-pătrat pentru un anumit tabel de contingență 2 x 2 , știm cât de probabil este să fi putut obține o valoare la fel de mare ca cea pe care am obținut-o de fapt strict întâmplător, în ipoteza independenței este cea corectă, adică dacă cele două categorii de clasificare nu au fost legate între ele sau dacă ipoteza nulă ar fi adevărată. Valoarea particulară a chi-pătratului pe care o obținem pentru exemplul nostru se întâmplă să fie 13,94.

Din cunoștințele noastre despre distribuția valorilor chi-pătratului, știm că, dacă ipoteza noastră nulă este adevărată, adică dacă nu există nicio diferență de mortalitate între grupul de control și cel tratat, atunci probabilitatea ca să obținem o valoare a chi-pătratului la fel de mare sau mai mare decât 13,94 doar din întâmplare este foarte, foarte scăzută; de fapt această probabilitate este mai mică de .005. Deoarece nu este probabil să obținem o valoare atât de mare a chi-pătratului din întâmplare în ipoteza noastră nulă, *trebuie să fie că aceasta a apărut nu întâmplător, ci pentru că ipoteza noastră nulă este incorectă* . Prin urmare, respingem ipoteza nulă la nivelul de semnificație .005 și acceptăm ipoteza alternativă, adică concluzionăm că printre femeile cu infarct miocardic noul medicament reduce mortalitatea. Probabilitatea de a obține aceste rezultate numai întâmplător este mai mică de 5/1000 (.005). Prin urmare, probabilitatea de respingere a ipotezei nule, atunci când aceasta este de fapt adevărată (eroare de tip I) este mai mică de .005.

Probabilitățile de obținere a diferitelor valori ale chi-pătratului sunt - tabelate în majoritatea textelor de statistică standard, astfel încât procedura este de a calcula valoarea chi-pătratului și apoi de a o căuta în tabel pentru a determina dacă este sau nu semnificativă. Acea valoare a chi-pătratului care trebuie obținută din date pentru a fi semnificativă se numește *valoare critică.* Valoarea critică a chi-pătratului la nivelul de semnificație 0,05 pentru un tabel 2 x 2 este 3,84. Aceasta înseamnă că atunci când obținem o valoare de 3,84 *sau mai mare* dintr-un tabel 2 x 2, putem spune că există o diferență semnificativă între cele două grupuri. Anexa A oferă câteva valori critice pentru chi-pătrat și pentru alte teste.

În utilizarea reală, se aplică o corecție pentru tabele 2 x 2 cunoscute sub numele de corecție Yates și calculul se face folosind formula:

$$\frac{|}{} \quad -|2N\_$$

$$\frac{N \left| I\, ad - bc\, I \right|^{-2}}{(a + b)\,(c + d)\,(a + c)\,(b + d)}$$

Notă: I *ad - bc* I înseamnă valoarea absolută a diferenței dintre *a x d* și *b x c*. Cu alte cuvinte, dacă *a x d* este mai mare decât *b x c,* scădem *bc* din *ad* ; dacă *bc* este mai mare decât *ad* , scade *ad* din *bc* . Chi-pătratul corectat astfel calculat este 12,95, încă mult peste 3,84 necesar pentru semnificație.

Testul chi-pătrat nu trebuie utilizat dacă numerele din celule sunt prea mici. Regulile generale: Când *N totalul* este mai mare de 40, utilizați testul chi-pătrat cu corecția lui Yates. Când *N* este între 20 și 40 și frecvența așteptată în fiecare dintre cele patru celule este de 5 sau mai mult, utilizați testul chi-pătrat corectat. Dacă cea mai mică frecvență așteptată este mai mică de 5 sau dacă *N* este mai mică de 20, utilizați testul Fisher.

În timp ce testul chi-pătrat aproximează probabilitatea, testul exact al lui Fisher oferă probabilitatea exactă de a obține un tabel cu valori ca cele obținute sau chiar mai extreme. Un exemplu de calcul este prezentat în Anexa B. Calculele sunt dificile, dar testul exact al lui Fisher este, de obicei, inclus în majoritatea programelor de statistică pentru calculatoare personale. Mai multe despre acest subiect pot fi găsite în cartea *Statistical Methods for Rates and Proportions* de Joseph L. Fleiss. Important este să știi când testul chi-pătrat este sau nu potrivit.

3.2 Testul McNemar

Să presupunem că avem situația în care se fac măsurători pe același grup de oameni înainte și după o intervenție sau să presupunem că suntem interesați de acordul dintre doi judecători care evaluează același grup de pacienți asupra unor caracteristici. În astfel de situații, măsurile înainte și după, sau opiniile a doi judecători, nu sunt independente unul față de celălalt, deoarece aparțin aceloraşi indivizi. Prin urmare, testul Chi-Pătrat sau Testul exact al lui Fisher nu sunt adecvate . În schimb, putem folosi testul McNemar.

Luați în considerare următorul exemplu. Cazurile pacienților care au fost suspectați de boală cardiacă ischemică (o scădere a fluxului sanguin către inimă din cauza înfundarii arterelor), au fost prezentate la doi experți în cardiologie. Medicii au fost rugați să își dea o opinie pe baza informațiilor disponibile despre pacient. Ei ar putea recomanda fie (1) ca pacientul să urmeze terapie medicală, fie (2) ca pacientul să aibă o angiografie, care este un test invaziv, pentru a determina dacă pacientul este un candidat potrivit pentru intervenția chirurgicală de bypass coronarian (cunoscut sub numele de CABG). Tabelul 3.1 prezintă rezultatele acestor judecăți pe 661 de pacienți.

TABELUL 3.1

		EXPERT 1		
		Medical	Chirurgical	
EXPERT 2	Medical	$a = 397$	$b = 97$	$a + b = 494$
	Chirurgical	$c = 91$	$d = 76$	$c + d = 167$
		$a + c = 488$	$b + d = 173$	$N = 661$

Rețineți că în celula *b* Expertul 1 a sfătuit intervenția chirurgicală și Expertul 2 a sfătuit terapie medicală pentru 97 de pacienți, în timp ce în celula *c* Expertul 1 a sfătuit terapie medicală și Expertul 2 a sfătuit intervenția chirurgicală pentru 91 dintre pacienți. Astfel, cei doi medici nu sunt de acord în 188 din cele 661 de cazuri sau 28% din timp. Celulele *a* și *d* reprezintă pacienți despre care cei doi medici sunt de acord. Ei sunt de acord în 473 din 661 de cazuri sau 72% din timp.

Pentru a determina dacă dezacordul observat ar fi putut apărea doar întâmplător sub ipoteza nulă a lipsei unui dezacord real în recomandările celor doi experți, calculăm un tip de valoare chi pătrat după cum urmează:

$$X^2 \text{ (chi-pătrat)} = \frac{. (lb - Cl - 1)^2}{} \text{---------} \text{---- - ----} \quad \frac{25}{188} = .13$$

(| *b* - *c* | înseamnă valoarea absolută a diferenței dintre cele două celule, adică indiferent de semn; -1 din numărător este analog cu corecția lui Yates pentru chi-pătrat descrisă în secțiunea 3.1 și oferă o mai bună aproximare a distribuției chi-pătrat.) Un chi pătrat de 0,13 nu atinge valoarea critică necesară pentru chi-pătrat 3.805. nivel de semnificație, așa cum este descris în Secțiunea 3.1, deci nu putem respinge ipoteza nulă și concluzionăm că datele noastre sunt în concordanță fără nicio diferență în opiniile celor doi experți. Dacă testul chi-pătrat semnificativ, ar trebui să respingem ipoteza nulă și să spunem că experții nu sunt în mod semnificativ de acord. Cu toate acestea, un astfel de test nu ne spune despre *puterea* acordului lor, care poate fi evaluat printr-o statistică numită Kappa.

3.3 Kappa

Cei doi experți ar putea fi de acord doar întâmplător, deoarece ambii experți sunt mai predispuși să recomande terapie medicală pentru acești pacienți. Kappa este o statistică care ne spune amploarea acordului dintre cei doi experți dincolo de acordul întâmplător.

$$\_ \frac{\textit{Proporția acordului observat - Proporția acordului întâmplător}}{1 - \textit{Proporția acordului întâmplător}}$$

Pentru a calcula numărul așteptat de cazuri în fiecare celulă a tabelului, urmează procedura descrisă pentru chi-pătrat în Secțiunea 3.1. Celulele *a* și *d* din tabelul 3.1 reprezintă acordul. Numărul așteptat doar întâmplător este

$$\textit{cell a}: \quad \frac{494 \times 488}{661} = 365$$

$$\textit{celula d}: \quad \text{----------} = 44 \quad \frac{167 \times 173}{661}$$

Deci proporția de acord așteptată doar din întâmplare este

$$\frac{365 + 44}{661} = .619$$

adică, numai întâmplător, experții ar fi de așteptat să fie de acord în 62% din cazuri. Proporția acordului observat este

$$661 \qquad \frac{397 + 76}{} = .716$$

$$\textit{Kappa} = \frac{.716 - .619}{1 - .619} = \frac{.097}{.381} = .25$$

Dacă cei doi experți ar fi de acord doar la nivel de șansă, Kappa ar fi 0; dacă cei doi experți ar fi de acord perfect, Kappa ar fi 1. Subiectul Kappa este descris în detaliu în cartea lui Fleiss enumerată în Lecturile sugerate.

3.4 Descrierea unei populații: Utilizarea abaterii standard

În cazul variabilelor continue, ca și în cazul variabilelor discrete, putem fi interesați de descriere sau de inferență. Când dorim să descriem o populație cu privire la o anumită caracteristică, folosim în general media sau media ca indice de *tendință centrală* a datelor.

Alte măsuri de tendință centrală sunt *mediana* și *modul* . Mediana este acea valoare peste care se află 50% din celelalte valori și sub care se află 50% din valori. Este valoarea medie sau a 50-a percentila. Pentru a găsi

mediana unui set de scoruri le aranjam în ordine crescătoare (sau descrescătoare) și găsim valoarea mijlocie dacă există un număr impar de scoruri, sau media dintre cele două scoruri mijlocii dacă există un număr par de scoruri. Modul este valoarea care apare cu cea mai mare frecvență. Pot exista mai multe moduri într-un set de scoruri, dar o singură valoare mediană și o valoare medie. Aceste definiții sunt ilustrate mai jos. Media este măsura ten denței centrale cel mai des folosită în statisticile inferențiale.

Măsuri de tendință centrală

Set de scoruri	Ordonat
12	6
12	8
6	10
8	**11 Mediană**
11	***12 Modul***
10	***12***
15	15
SUMA: 74	*Medie* = 74/7 = 10,6

Adevărata medie a populației se numește *m* și o estimăm din datele obținute dintr-un eșantion al populației. Media eșantionului se numește *x* (se citește ca x bară). Trebuie să avem grijă să precizăm exact populația din care prelevăm un eșantion. De exemplu, în populația generală, IQ-ul mediu este de 100, dar IQ-ul mediu al populației de copii cu vârsta cuprinsă între 6 și 11 ani ai căror tați sunt absolvenți de facultate este de 112,9. Prin urmare, dacă luăm un eșantion din oricare dintre aceste populații, am estima o medie a populației diferită și trebuie să specificăm pentru care populație facem inferențe.

Cu toate acestea, media nu oferă o descriere adecvată a unei populații. De asemenea, este nevoie de o măsură a *variabilității* datelor în jurul valorii medii. Două grupuri pot avea aceeași medie, dar pot fi foarte diferite. De exemplu, luați în considerare un grup ipotetic de copii al căror IQ individual

este 100; astfel, media este 100. Comparați aceasta cu un alt grup a cărui medie este tot 100, dar include indivizi cu IQ-ul este de 60 și cei cu IQ-ul de 140. Trebuie făcute diferite afirmații despre aceste două grupuri: unul este compus din toți indivizii medii; celălalt include atât retardați, cât și genii.

Cel mai frecvent utilizat indice de variabilitate este *deviația standard (sd)* , care este un tip de măsură legat de distanța medie a scorurilor față de valoarea lor medie. Pătratul abaterii standard se numește *varianță* . Abaterea standard a populației este notată cu litera greacă c (sigma). Când este calculat dintr-un *eșantion,* este scris ca sd și este ilustrat în exemplul de mai jos:

Scoruri IQ		**Abateri de la medie**		**Scoruri la pătrat pentru B**
Grupa A	Grupa B	$X_{i} - x_{B}$	$(Xi - XB)^2$	X_B^2
100	60	-40	1600	3600
100	140	40	1600	19.600
100	80	-20	400	6400
100	120	20	400	14.400
S = 400 X_A = medie = 100	**S** = 400 X_B = medie = 100	**S** = 0	**S** = 4000 de abateri la pătrat	**S** = 44.000 suma pătratelor

Notă: simbolul „ S" înseamnă „suma".

Notă: Suma abaterilor de la medie, ca în coloana 3, este întotdeauna 0; de aceea însumăm abaterile pătrate, ca în coloana 4.

$\bar{x}_A$ = *medie* = $\frac{400}{4}$ = 100; $\bar{x}_B$ = $\frac{400}{4}$ = 100

$$sd = \sqrt{\frac{\sum (\text{din (fiecare valoare - media grupului)}^2}{n - 1}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

$$sd_A = \sqrt{\frac{0}{3}} = 0;$$

(În grupa A, deoarece fiecare scor este egal cu media lui 100, nu există abateri de la media lui A)

$$sd_B = \sqrt{\frac{4000}{3}} = \sqrt{1333} = 36.51$$

O formulă echivalentă pentru sd care este mai potrivită pentru calculele reale este

$$sd = \sqrt{\frac{\sum X_i^2 - n\bar{X}^2}{n - 1}}$$

Pentru grupa B avem

$$sd = \sqrt{\frac{44000 - 4(100)^2}{4} \cdot \frac{1}{3}} = \sqrt{\frac{44000 - 40000}{3} \cdot \frac{1}{4000}} = 36.51$$

Varianta = (*sd*)2

Rețineți că media ambelor grupuri este 100, dar abaterea standard a grupului A este 0, în timp ce sd a grupului B este 36,51. (Împărțim abaterile la pătrat la *n* - 1, mai degrabă decât la *n,* deoarece estimăm populația *G* din datele eșantionului, iar împărțirea la *n* - 1 oferă o estimare mai bună. Motivul matematic este complex și dincolo de scopul acestei cărți.)

3.5 Semnificația abaterii standard: distribuția normală

Abaterea standard este o măsură a dispersiei sau răspândirii datelor. Luați în considerare o variabilă precum IQ, care este distribuită în mod normal, adică poate fi descrisă prin curba familiară, în formă de clopot, în care majoritatea valorilor se încadrează în jurul mediei cu un număr descrescător de valori la ambele extreme. Într-un astfel de caz, 68% dintre valori se află în limita unei abateri standard de fiecare parte a mediei, 95% dintre valori se află la 2 abateri standard ale mediei și 99% dintre valori se află la 3 abateri standard ale mediei.

Acest lucru este ilustrat în Figura 3.1.

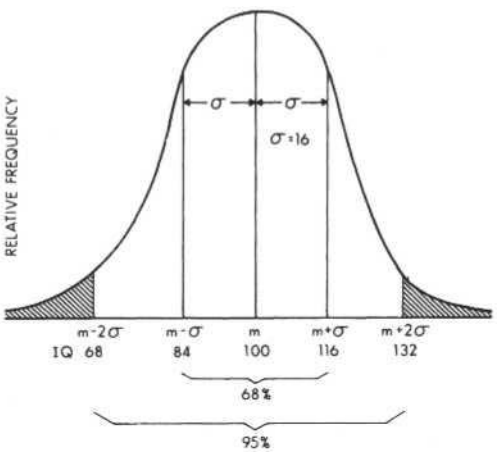


Figura 3.1

În populația în general, 95% dintre oameni au IQ-uri între 68 și 132. Aproximativ 2,5% dintre oameni au IQ-uri peste 132 și alte 2,5% dintre oameni au IQ-uri sub 68. (Acest lucru este indicat de zonele umbrite de la

cozile curbelor.)

Dacă estimăm dintr-un eșantion și dacă există un număr mare de observații, abaterea standard poate fi estimată din *intervalul datelor,* adică diferența dintre valoarea cea mai mică și cea mai mare. Împărțirea intervalului la 6 oferă o estimare aproximativă a abaterii standard dacă distribuția este normală, deoarece 6 abateri standard (3 de fiecare parte a mediei) cuprind 99% sau aproape toate datele.

La nivel individual, clinic, cunoașterea abaterii standard este foarte utilă pentru a decide dacă o constatare de laborator este normală, în sensul „sănătos". În general, o valoare care este la mai mult de 2 deviații standard de la medie este suspectă și poate că trebuie efectuate teste suplimentare.

De exemplu, să presupunem că, ca medic, vă confruntați cu un bărbat adult care are o valoare a hematocritului de 39. Hematocritul este o măsură a cantității de globule roșii într-o cantitate măsurată de sânge. Un hematocrit scăzut poate implica anemie, care la rândul său poate implica o afecțiune mai gravă. De asemenea, știți că valoarea medie a hematocritului la bărbați adulți este de 47. Știți dacă pacientul cu o valoare de 39 este normală (în sensul de sănătos) sau anormal? Trebuie să cunoașteți abaterea standard a distribuției hematocritelor la oameni înainte de a putea determina dacă 39 este o valoare normală. De fapt, abaterea standard este de aproximativ 3,5; astfel, plus sau minus 2 abateri standard în jurul mediei rezultă în intervalul de la 40 la 54, astfel încât 39 ar fi ușor scăzut. Pentru femelele adulte, hematocritul mediu este de 42 cu o abatere standard de 2,5, astfel încât intervalul de plus sau minus 2 abateri standard de la medie este de la 37 la 47. Astfel, dacă o femeie adultă vine la tine cu o citire a hematocritului de 39, ar fi considerată în intervalul „normal".

3.6 Diferența dintre abaterea standard și eroarea standard

Adesea datele din literatură sunt raportate ca ± sd (citite ca medie + sau -1 abatere standard). Alteori sunt raportate ca ± se (se citesc ca medie + sau - 1 eroare standard). *Eroarea standard* și *abaterea standard* sunt adesea confundate, dar au funcții destul de diferite. Pentru a înțelege conceptul de eroare standard, trebuie să vă amintiți că scopul statisticilor este de a trage inferențe din eșantioane de date către populația din care provin aceste eșantioane. Mai exact, ne interesează estimarea adevăratei medii a unei populații pentru care avem o medie eșantională bazată, de exemplu, pe 25 de cazuri. Imaginați-vă următoarele:

Populația **Scoruri IQ,** x_i	**Eșantion înseamnă bazat pe** **25 de persoane selectate aleatoriu**
110	$X = 102$
100	
105	~ = 99
98	
140	$X_3 = 101$ $x_4 = 98$
100	100
m = mean of all the x_i 's	$m_x = m$ media a mijloacelor este m , media populației
σ = population standard deviation	abaterea standard a distri butiunea X -urilor numită stan eroare dard a mediei = $\frac{\sigma}{\sqrt{n}}$

Există o populație de scoruri IQ a cărei medie este 100 și abaterea sa standard este 16. Acum imaginați-vă că extragem la întâmplare un eșantion de 25 de persoane din acea populație și calculăm media eșantionului

x. Această medie a eșantionului se întâmplă să fie 102. Dacă am lua un alt eșantion de 25 de indivizi, probabil că am obține o medie a eșantionului ușor diferită, de exemplu 99. Să presupunem că am făcut acest lucru în mod repetat de un număr infinit (sau de un număr foarte mare) de ori, de fiecare dată când aruncăm eșantionul pe care tocmai l-am extras înapoi în populația din care am eșantionat din nou 25 de persoane. Am avea atunci un număr foarte mare de astfel de mijloace de eșantionare. Aceste mijloace eșantionului ar forma o distribuție normală. Unele dintre ele ar fi foarte aproape de media reală a populației de 100, iar unele s-ar afla la fiecare capăt al acestei „distribuții a mijloacelor", ca în figura 3.2.

Această distribuție a mijloacelor eșantionului ar avea propria sa deviație standard , adică o măsură a răspândirii datelor în jurul valorii medii a datelor. În acest caz, datele sunt medii eșantionare mai degrabă decât valori individuale. Abaterea standard a acestei distribuții a mediilor se numește *eroarea standard a mediei* .

Trebuie subliniat că această distribuție a mijloacelor, numită și distribuția prin eșantionare a mijloacelor, este o construcție teoretică. Evident, nu ocolim măsurarea eșantioanelor din populație pentru a construi o astfel de distribuție. De obicei, de fapt, luăm doar *un eșantion de 25* de persoane și ne imaginăm ce ar putea această distribuție
b e. Cu toate acestea, datorită anumitor derivări matematice, știm multe despre această distribuție teoretică a mediilor populației și, prin urmare, putem trage inferențe importante pe baza unei singure medii de eșantion. Ceea ce știm este că distribuția mediilor este o distribuție normală , că media ei este aceeași cu media populației valorilor individuale, adică *media mediilor este m,* și că abaterea sa standard este egală cu abaterea standard a valorilor individuale originale împărțită la rădăcina pătrată a numărului de persoane din eșantion.

Eroarea standard a mediei =

$$O = \frac{o}{T}$$

In acest caz ar fi

$$\frac{\frac{16}{725}}{\frac{16}{5}} = 3.2$$

Distribuția mijloacelor ar arăta așa cum se arată în Figura 3.2.

Vă rugăm să rețineți că atunci când vorbim despre valorile populației, pe care de obicei nu le cunoaștem, dar încercăm să le estimăm, ne referim la medie ca *m* și abaterea standard ca *a* . Când vorbim despre valorile calculate din probe, ne referim la medie ca *x* , abaterea standard ca sd și eroarea standard ca se

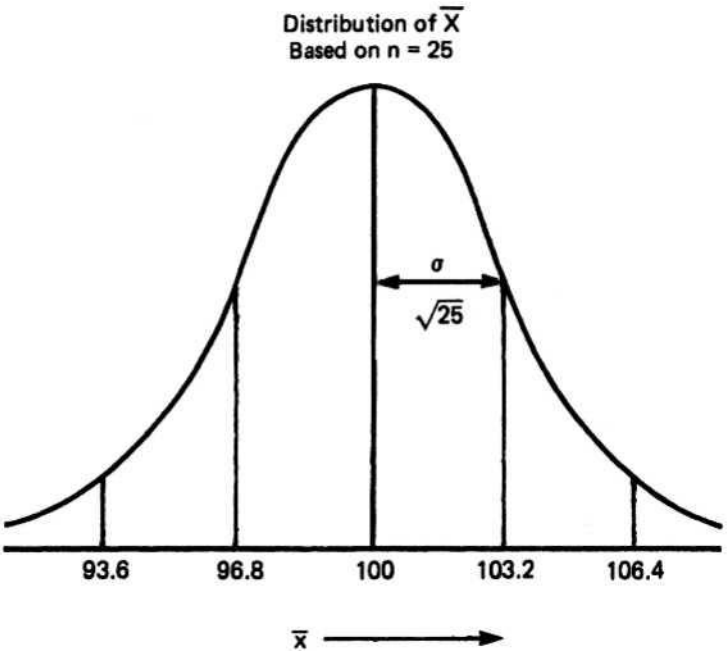


Figure 3.2

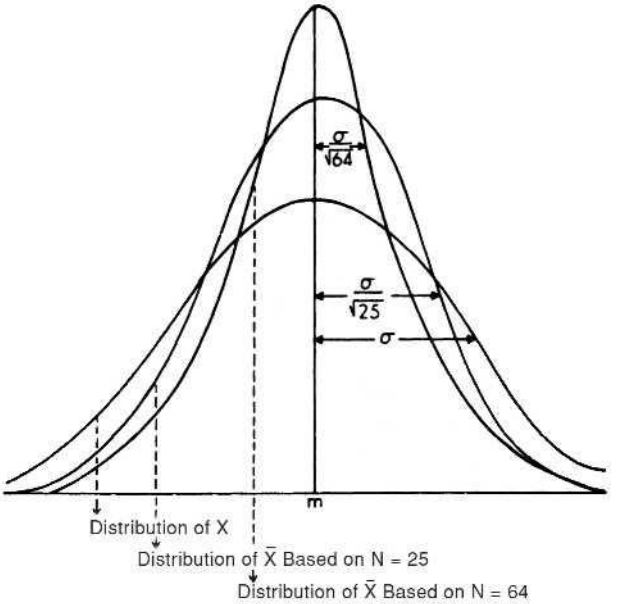


Figura 3.3

Acum imaginați-vă că avem o distribuție a mijloacelor bazată pe eșantioane de 64 de indivizi. Media acestor mijloace este, de asemenea , *m,* dar dispersia sa , sau eroarea standard, este mai mică. Este 16 / -764 = 16 / 8 = 2. Acest lucru este ilustrat în Figura 3.3.

Se poate observa cu ușurință că, dacă luăm un eșantion de 25 de indivizi, este probabil ca media lor să fie mai aproape de media adevărată decât valoarea unui singur individ, iar dacă extragem un eșantion de 64 de indivizi, este probabil ca media lor să fie chiar mai aproape de media adevărată decât a fost media pe care am obținut-o din eșantionul de 25.

Abaterea standard este utilizată pentru a descrie dispersia sau variabilitatea scorurilor. Eroarea standard este utilizată pentru a deduce inferențe despre media populației din care avem un eșantion. Tragem astfel de inferențe construind intervale de încredere, care sunt discutate în Secțiunea 3.11.

3.7 Eroare standard a diferenței dintre două mijloace

Acest concept este analog cu conceptul de eroare standard a mediei. Eroarea standard a diferențelor dintre două medii este abaterea standard a unei distribuții teoretice a diferențelor între două medii. Imaginați-vă un grup de bărbați și un grup de femei, fiecare având o măsurătoare IQ. Să presupunem că luăm un eșantion de 64 de bărbați și un eșantion de 64 de femei, calculăm IQ-urile medii ale acestor două eșantioane și obținem diferențele lor. Dacă ar fi să facem acest lucru de un număr infinit de ori, am obține o *distribuție a diferențelor* între mediile eșantionului a două grupuri de 64 fiecare. Aceste scoruri de diferență ar fi distribuite în mod normal; media lor ar fi adevărata diferență medie dintre populațiile de bărbați și femei (pe care încercăm să o deducem din eșantioane), iar abaterea standard a acestei distribuții se numește *eroarea standard a diferențelor dintre două medii.*

Eroarea standard a diferenței dintre două medii ale populațiilor X *și* Y *este* dată de formula

$$Q_{x-y} \qquad \frac{1}{y\,nx} \qquad \frac{Q^{2}_{y}}{n_y}$$

$$22$$

unde Q_x este varianța populației X și Q_y este varianța populației Y; n_x este numărul de cazuri din eșantionul din populația X și n_y este numărul de cazuri din eșantionul din populația Y.

În unele cazuri, știm sau presupunem că variațiile celor două populații sunt egale între ele și că variațiile pe care le calculăm din eșantioanele pe care le-am extras sunt ambele estimări ale unei variații comune. Într-o astfel de situație, am dori să punem în comun aceste estimări pentru a obține o estimare mai bună a varianței comune. Notăm această *estimare combinată* ca $s^2_{pooled} = s_p{}^2$ și calculăm eroarea standard a diferenței dintre medii ca

$$\sigma_{x-y} \quad \sqrt{s_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y}\right)} \qquad s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

Calculăm s_p din datele eșantionului:

$$s_p^2 \quad \frac{(n_x - 1) s_x^2 + (n_y - 1) s_y^2}{n_x + n_y - 2}$$

Acesta este echivalentul cu

$$\frac{\sum (x - \bar{x})^2 + \sum (y_t - \bar{y})^2}{n_x + n_y - 2}$$

Deoarece, în practică, vom calcula întotdeauna valorile noastre din datele eșantionului, vom folosi de acum înainte simbolologia adecvată.

3.8 Scorurile Z și distribuția normală standardizată

Distribuția normală standardizată este una a cărei medie = 0, abaterea standard = 1 și aria totală sub curbă = 1. Distribuția normală standard arată ca în Figura 3.4.

Pe abscisă, în loc de *x* avem o transformare a lui *x* numită scor standard, Z. Z este derivat din *x* prin următoarele:

$$z = \frac{x - m}{a}$$

Astfel, scorul Z vă spune cu adevărat câte abateri standard de la medie este un anumit scor *x* .

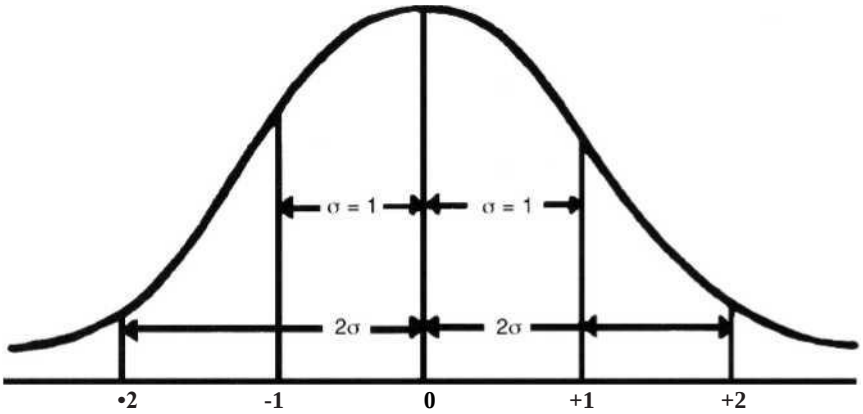


Figura 3.4

Orice distribuție a unei variabile normale poate fi transformată într-o distribuție a lui Z luând fiecare valoare *x* , scăzând din ea media lui *x* (adică, *m)* și împărțind această abatere a lui *x* de la media ei, la abaterea standard. Să ne uităm din nou la distribuția IQ în Figura 3.5.

Astfel, un scor IQ de 131 este echivalent cu un scor Z de 1,96 (adică este 1,96, sau aproape 2, abateri standard peste IQ-ul mediu).

$$z = \frac{131 - 100}{16} = L96$$

Unul dintre lucrurile frumoase despre distribuția Z este că probabilitatea ca o valoare să se afle oriunde între două puncte este egală cu aria de sub curba dintre aceste două puncte. (Acceptați acest lucru pe credință.) Se întâmplă ca aria din dreapta lui 1,96 să corespundă unei probabilități de .025, sau 2,5% din curba totală. Deoarece curba este simetrică, probabilitatea ca Z să fie la stânga lui -1,96 este de asemenea .025. Invocând ad-

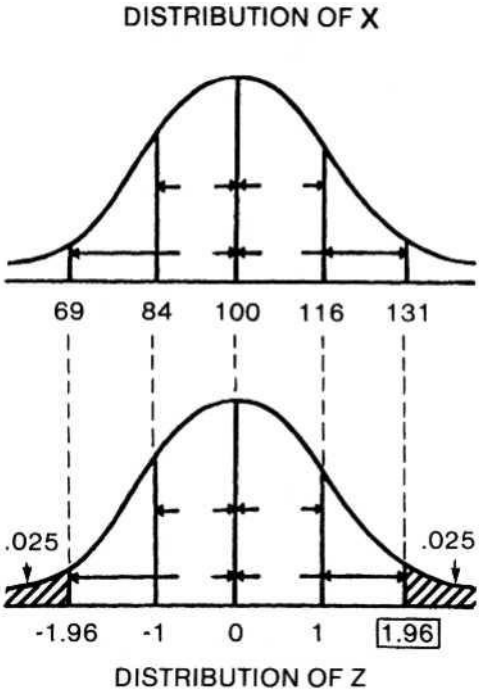


Figura 3.5

legea ditivă a probabilității (Secțiunea 2.2), probabilitatea ca un Z să fie *fie* la stânga lui -1,96 , *fie* la dreapta lui +1,96 este .025 + .025 = .05. Transformându -ne înapoi la *x* , putem spune că probabilitatea ca cineva să aibă un IQ în afara 1,96 deviații standard de la medie (adică peste 131 sau sub 69) este 0,05, sau doar 5% din populație are valori atât de extreme. (De obicei, valoarea Z de 1,96 este rotunjită la ± 2 abateri standard de la medie, corespunzătoare punctelor limită dincolo de care se află 5% din curbă, dar valoarea exactă este 1,96.)

O utilizare foarte importantă a lui Z derivă din faptul că putem converti, de asemenea, o medie a eșantionului (mai degrabă decât o singură valoare individuală) într-un scor Z.

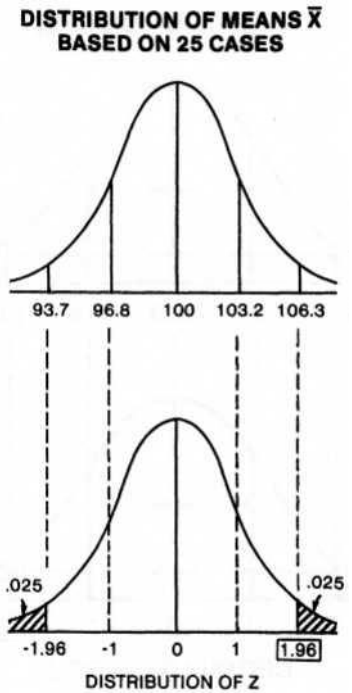


Figure 3.6

$$7\_\ \frac{x - m}{G\sim_X}$$

Numătorul acum este distanța dintre media eșantionului de media populației, iar numitorul este abaterea standard a distribuției mediilor, care este *eroarea standard a mediei* . Acest lucru este ilustrat în Figura 3.6, unde luăm în considerare mijloace bazate pe 25 de cazuri fiecare. Se este 16 / V25 = 16 / 5 = 3 . 2.

Acum putem vedea că o medie a eșantionului de 106,3 corespunde unui scor Z de 1,96.

$$Z = \frac{106\ 3 - 100}{3.2} = {}_{1>96}$$

Putem spune acum că probabilitatea ca *IQ-ul mediu al unui grup de 25 de*

persoane să fie mai mare decât 106,3 este .025. Probabilitatea ca o astfel de medie să fie mai mică de 93,7 este, de asemenea, de .025.

Scorul AZ poate fi calculat și pentru *diferența dintre două medii.*

$$Z = \frac{(X_A - X_B) - (m_A - m_B)}{\sigma_{X_A - X_B}}$$

Dar $m_A - m_B$ este în mod obișnuit ipotezat a fi 0, astfel încât formula be vine

$$\frac{X_A - X_B}{\sigma_{X_A - X_B}}$$

Puteți vedea că un *scor Z în general este o distanță între o anumită valoare și media ei împărțită la o eroare standard adecvată* .

Acest lucru devine foarte util mai târziu când vorbim despre intervalele de încredere în secțiunile 3.10 până la 3.14.

3.9 Statistica t

Să presupunem că suntem interesați de mediile eșantionului și dorim să calculăm un scor Z. Nu știm care este abaterea standard a populației, dar dacă eșantioanele noastre sunt foarte mari, putem obține o estimare bună a lui σ prin calculul abaterii standard, sd, din eșantionul nostru și apoi obținem eroarea standard ca de obicei: se = sd/Tn. Dar adesea eșantionul nostru nu este suficient de mare. Putem obține în continuare un scor standardizat calculând o valoare numită t al lui Student:

$$t = \frac{xm}{se.x}$$

Arată exact ca Z; singura diferență este că o calculăm din probă și este o probă mică.

Putem obține probabilitatea de a obține anumite valori t în mod similar modului în care am obținut probabilitățile de valori Z - dintr-un tabel adecvat. Dar se întâmplă că, în timp ce distribuția t arată ca o distribuție Z normală, este doar puțin diferită, dând astfel probabilități ușor diferite. De fapt, există multe distribuții t (nu doar una, ca pentru Z). Există o distribuție t diferită pentru fiecare dimensiune diferită a eșantionului. (Vor fi explicate mai multe despre acest lucru în Secțiunea 3.10.)

În exemplul nostru, unde avem o medie bazată pe 25 de cazuri, am avea nevoie de o valoare de 2,06 pentru a corespunde unei probabilități de 0,025 (în loc de 1,96 pentru distribuția Z). Traducând acest lucru înapoi la scara eșantionului, dacă eroarea noastră standard ar fi 3,2, atunci probabilitatea ar fi 0,025 că am obține o medie a eșantionului de 106,6 (care este 100 + 2,06 ori 3,2), mai degrabă decât 106,3 (care este de 100 + 1,2,96 ori în distribuția Z). Acest lucru poate părea o slăbiciune, deoarece diferențele sunt atât de mici. De fapt, pe măsură ce dimensiunea eșantionului se apropie în infinit, distribuția t devine exact ca distribuția Z, dar diferențele dintre Z și t devin mai mari pe măsură ce dimensiunea eșantionului devine mai mică și este întotdeauna sigur să utilizați distribuția t. De exemplu, pentru o medie bazată pe cinci cazuri, valoarea t ar fi 2,78 în loc de Z de 1,96. Unele valori t sunt prezentate în Anexa A. Tabelele mai detaliate sunt în cărțile standard de statistică.

3.10 Valori eșantioane și valori ale populației revizuite

Toate acestea mergând înainte și înapoi între valorile eșantionului și valorile populației pot fi confuze. Iată punctele de reținut:

(1) Suntem întotdeauna interesați de estimarea valorilor populației din eșantioane.
(2) În unele dintre formule și termeni, folosim valorile populației ca și cum am ști care sunt cu adevărat valorile populației. Desigur

, nu cunoaștem valorile efective ale populației, dar dacă avem eșantioane foarte mari, le putem estima destul de bine din datele eșantionului nostru.

(3) În scopuri practice, în general, vom folosi și ne vom referi la tehnici adecvate pentru eșantioane mici, deoarece acest lucru este mai comun și mai sigur (adică, nu doare chiar dacă avem mostre mari).

3.11 O întrebare de încredere

Un interval de încredere stabilește un interval și specifică probabilitatea ca acest interval să cuprindă media reală a populației. De exemplu, un interval de încredere de 95% (aproximativ) este stabilit prin luarea mediei eșantionului, x , plus sau minus *două erori standard ale mediei.*

Interval de încredere de 95%:

$$x \pm 2\ se = x \pm 2\ \hat{}4$$

Astfel, dacă am luat un eșantion aleatoriu de 64 de solicitanți la Colegiul de Medicină Albert Einstein și am constatat că IQ-ul lor mediu este de 125, să zicem, (o cifră fictivă), am dori să stabilim un interval de încredere de 95% pentru a deduce care este cu adevărat media reală a populației de solicitanți. Intervalul de încredere de 95% este intervalul cuprins între 125-2 se și 125 + 2s.e. De obicei exprimăm asta ca,

> „Suntem 95% încrezători că adevăratul IQ mediu al candidaților la facultatea de medicină Einstein se află în 125 ± 2 se."

În sensul acestui exemplu, presupunem că abaterea standard este 16. (Acesta nu este o ipoteză deosebit de bună, deoarece variabilitatea IQ-ului solicitanților de la facultatea de medicină este considerabil mai mică decât variabilitatea IQ-ului în populație în general.) În baza acestei ipoteze , ajungem la următorul interval:

$$125 \pm \frac{}{764} = 125 \pm \frac{}{8} = 125 \pm 4 = 121\text{-}129$$

Declarația noastră acum este următoarea: „Probabilitatea este de aproximativ .95 ca adevăratul IQ mediu al solicitanților de la Einstein Medical School să se situeze în intervalul 121-129." (O interpretare mai riguroasă a acestui lucru este dată în Secțiunea 3.11.)

Un interval de încredere de 99% este aproximativ media eșantionului ± 3s.e. În exemplul nostru, acest interval ar fi:

$$125 \pm 3 \frac{(16)}{\bullet 764} = 125 \pm 6 = 119\text{-}131$$

Atunci am putea spune: „Probabilitatea este de aproximativ .99 ca adevăratul IQ mediu al solicitanților de la Einstein Medical School să se situeze în intervalul 119-131."

„Aproximativ" se datorează faptului că pentru a obține o probabilitate de .95 nu înmulți se-ul cu 2 exact așa cum am făcut aici; l-am rotunjit pentru comoditate . Factorul de multiplicare *exact* depinde de cât de mare este eșantionul. Dacă eșantionul este foarte mare, mai mare de 100, am înmulți se cu 1,96 pentru intervale de încredere de 95% și cu 2,58 pentru intervale de încredere de 99%. Dacă eșantionul este mai mic, ar trebui să căutăm multiplicatorul în tabelele cu valorile t, care apar în multe texte. Aceste valori t sunt diferite pentru diferite „grade de libertate", explicate în Secțiunea 3.13, care sunt legate de dimensiunile eșantionului. Unele valori t sunt prezentate în Anexa A. (Consultați, de asemenea, înapoi la Secțiunea 3.9 pentru semnificația statisticilor t .)

Rețineți că pentru o anumită dimensiune a eșantionului schimbăm gradul de certitudine cu dimensiunea intervalului. Putem fi mai siguri că adevărata noastră medie se află într-un interval mai larg, dar dacă dorim să stabilim intervalul mai precis , suntem mai puțin siguri despre ea (Figura 3.7). Pentru a obține mai multă precizie și a menține o probabilitate ridicată de a fi corect în estimarea

INTERVALE DE ÎNCREDERE PENTRU IQ

95% ÎNCREDERE

Figura 3.7

interval, este necesar să măriți dimensiunea eșantionului. Principalul punct este că atunci când raportați o medie a eșantionului ca estimare a mediei unei populații, este cel mai de dorit să raportați limitele de încredere.

3.12 Limite de încredere și intervale de încredere

Limitele de încredere sunt granițele exterioare pe care le calculăm și despre care putem spune: suntem 95% încrezători că aceste limite sau limite includ adevărata medie a populației. Intervalul dintre aceste limite se numește *interval de încredere.* Dacă ar fi să luăm un număr mare de eșantioane din populație și să calculăm limitele de încredere de 95% pentru fiecare dintre ele, 95% din intervalele limitate de aceste limite ar conține media reală a populației. Cu toate acestea, 5% nu l-ar conține. Desigur, în viața reală luăm doar un eșantion și construim intervale de încredere din acesta. Nu putem fi niciodată siguri dacă intervalul calculat din eșantionul nostru particular este unul dintre cei 5% dintre astfel de intervale care nu conțin media populației. Cel mai mult putem spune că suntem 95% încrezători că îl conține. După cum puteți vedea, nu știm niciodată nimic sigur.

Dacă din populația de interes au fost extrase un număr infinit de eșantioane aleatoare independente (cu înlocuire), atunci 95% din intervalele de încredere calculate din eșantioane (cu media x și eroarea standard se) vor cuprinde adevărata medie a populației m.

Figura 3.8 ilustrează conceptele de mai sus.

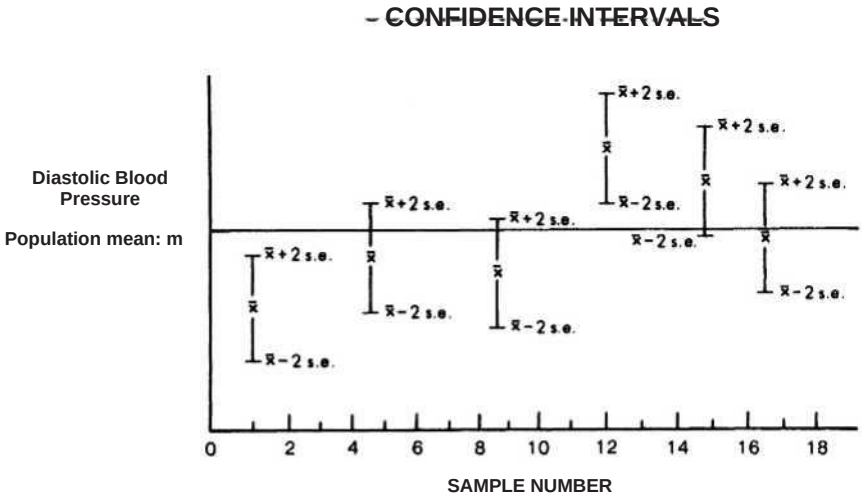


Figure 3.8

3.13 Grade de libertate

Valorile t pe care le folosim ca multiplicator al erorii standard pentru a - construi intervale de încredere depind de ceva numit *grade de libertate* (df), care sunt legate de dimensiunea eșantionului. Când avem un eșantion, pentru a găsi valoarea t adecvată pentru a calcula limitele de încredere, introducem tabelele cu *n* - 1 grade de libertate, unde *n* este dimensiunea eșantionului. O modalitate intuitivă de a înțelege conceptul de df este să considerăm că, dacă calculăm media unui eșantion de, să zicem, trei valori, am avea „libertatea" de a varia două dintre ele în orice fel ne-ar plăcea după ce știm care este media, dar a treia trebuie fixată pentru a ajunge la media dată. Deci avem doar 2 „grade de libertate". De exemplu, dacă știm că media a trei valori este 7, putem avea următoarele seturi de date:

Valoarea 1:	7	-50
Valoarea 2:	7	+18
Valoarea 3:	7	+53
Suma =	21	21

Media = $v = 7$ $V = 7$

În fiecare caz, dacă cunoaștem valorile 1 și 2, atunci se determină valoarea 3 deoarece suma acestor valori trebuie să fie 21 pentru ca media să fie 7. Am „pierdut" un grad de libertate în calcularea mediei.

3.14 Intervale de încredere pentru proporții

O proporție poate fi considerată o variabilă continuă. De exemplu, în studiul anticoagulant descris în secțiunea 3.1, proporția femeilor din grupul de control (tratat cu placebo) care au supraviețuit unei inimi s-a dovedit a fi 89/129 = 0,69. O proporție poate lua valori de-a lungul continuumului între 0 și 1. Putem construi un interval de încredere în jurul unei proporții într-un mod similar cu construirea intervalelor de încredere în jurul mediilor. Limitele de încredere de 95% pentru o proporție sunt $p \pm 1{,}96$ se $_p$, unde se $_p$ este *eroarea standard a unei proporții.*

Pentru a calcula eroarea standard a unei proporții, trebuie mai întâi să calculăm abaterea standard a unei proporții și să o împărțim la rădăcina pătrată a lui n . Ne definim simbolologia:

s = abaterea standard a unei proporții = -Jpq

, . numărul de supraviețuitori din grupul de control
p = proporția eșantionului = ---------- - ---------- ^{2}____
numărul total de femei din grupul de control
numărul decese în grupul de control
numărul total de femei din grupul de control

$$s.e._p = \frac{\sqrt{pq}}{\sqrt{n}} = \sqrt{\frac{pq}{n}}$$

În exemplul nostru de femei care au supraviețuit unui atac de cord din grupul de control, intervalul de încredere de 95% este

$$.69 \pm 1{,}96\,x\,\frac{(-69)\,x\,(-31)}{V\ 129} = .69 \pm .08$$

Și putem afirma că suntem 95% încrezători că proporția populației de femei netratate care supraviețuiesc unui atac de cord este între .61 și .77 sau 61% și 77%. (Rețineți că aceasta se referă la populația din care a fost extras eșantionul nostru. Nu putem generaliza acest lucru la toate femeile care au un atac de cord.)

Pentru limitele de încredere de 99%, am înmulți eroarea standard a unei proporții cu 2,58, pentru a obține intervalul de la .59 la .80. Multiplicatorul este valoarea Z care corespunde cu 0,95 pentru limitele de încredere de 95% sau cu probabilitatea .99 pentru limitele de încredere de 99%.

3.15 Intervale de încredere în jurul diferenței dintre două mijloace

Putem construi intervale de încredere în jurul unei diferențe dintre medii într-un mod similar cu care am construit intervalele de încredere în jurul unei singure medii. Limitele de încredere de 95% în jurul diferenței dintre medii sunt date de

$$(x - {}^{y}) \pm (t_{d} \quad (Se_{xy})$$

În cuvinte, aceasta este diferența dintre cele două medii eșantionare, plus sau minus o valoare t adecvată, ori eroarea standard a diferenței; df este gradele de libertate și .95 spune că căutăm valoarea t care se referă la acele grade de libertate și la probabilitatea .95. Gradele de libertate atunci când ne uităm la două eșantioane sunt $n_x + n_y - 2$. Acest lucru se datorează faptului că am pierdut un grad de libertate pentru fiecare dintre cele două mijloace pe care le-am calculat, deci gradele noastre totale de libertate sunt $(n_x - 1) + (n_y - 1) = n_x + n_y - 2$.

Ca exemplu, luați în considerare că avem un eșantion de 25 de studenți la medicină și 25 de bărbați. IQ-urile medii pentru fiecare probă sunt

$$^{x}bărbați = 130, \; ^{x}bărbați = 126, \; Sp_{0\,0}l_{e}d = 12, \; _{df} = 48$$

Intervalul de încredere de 95% pentru diferența medie între bărbați și femei se calculează după cum urmează:

Din tabelele t, aflăm că valoarea t pentru df = 48 este 2,01

$$^{x}males \quad \pm 2.01 \times \sqrt{S_p \; \frac{1}{n_x} + \frac{1}{n_{y\,J}}}$$

$$(130 - 126) \pm 2 \,.\, 01 \times 712(1 / 25 + 1 / 25) = 4 \pm 6 \,.\, 8$$

Intervalul este atunci de la -2,8 la 10,8 și suntem 95% siguri că include adevărata diferență medie între bărbați și femei. Acest interval include 0 diferență, așa că ar trebui să concluzionam că diferența de IQ între bărbați și femei poate fi zero.

3.16 Comparații între două grupuri

Cea mai frecventă problemă care apare este necesitatea de a compara două grupuri pe o anumită dimensiune. Am putea dori să stabilim, de exemplu, dacă (1) administrarea unui anumit medicament scade tensiunea arterială sau (2) medicamentul A este mai eficient decât medicamentul B în scăderea nivelului zahărului din sânge sau (3) învățarea copiilor de clasa întâi să citească prin metoda I produce scoruri mai mari la sfârșitul anului decât învățarea lor să citească prin metoda II.

3.17 Testul Z pentru compararea a două proporții

Ca exemplu, reproducem aici tabelul din Secțiunea 3.1 care prezintă date dintr-un studiu privind terapia anticoagulantă.

Frecvențele observate

	Controla	Tratat	
A trăit	89	223	312
A murit	40	39	79
Total	129	262	391

Dacă dorim să testăm dacă proporția de femei care au supraviețuit unui atac de cord în grupul tratat diferă de proporția care au supraviețuit în grupul de control, am stabilit ipoteza noastră nulă ca

HO : $P_1 = P_2$ sau $P_1 - P_2 = 0$; P_1 = proporția supraviețuitoare în populația tratată

P_2 = proporția de supraviețuire în populația martor

H_A : $P_1 - P_2 \neq 0$ (diferența nu este egală cu 0)

We calculate

$$\frac{P_1 - P_2}{s.e_{p-P_2}}$$

P_i $\frac{223}{262}$ = .85, q_1 = 1 - P_i = .15, m = 262

P_2 $\frac{89}{129}$ = 1 - P_2 = .31, n_2 = 129

Astfel, numărătorul lui Z = .85 - .69 = .16.

Numitorul =
eroare standard a diferenței dintre două proporții =

$$\wedge.(P\text{-}P?) = J^{pq} \quad 7 \text{---} \quad \text{-------}\backslash \quad I^{n_1} \quad \frac{1}{n_2}$$

unde P și q sunt estimări combinate bazate pe date de grup. O calculăm după cum urmează:

on both treated and control

$$P = \frac{n_1 p_1 + n_2 p_2}{\text{---------- -- ----------}} = \frac{\textit{numărul de supraviețuitori în tratați + control}}{\text{---------- -- ----------}}$$

$n_1 + n_2$ *numărul total de pacienţi în trataţi + control*
$q = 1 - p = 1 - .80 = .20$

$$= \frac{262(.85) + 129(.69)}{262 + 129} = \frac{223 + 89}{391} = 80$$

$$s.e.lr,\_r,, = \wedge .80)(.20)\left[\frac{}{262 + 129}\right] = .043$$

$$z \_ \frac{.85 - .69}{.043} \quad 3.72$$

Trebuie să ne uităm acum pentru a vedea dacă această valoare a lui Z depășește *valoarea critică* . *Valoarea critică este valoarea minimă a statisticilor de test pe care trebuie să o obținem pentru a respinge ipoteza nulă la un anumit nivel de semnificație* .

Valoarea *critică a lui Z* pe care trebuie să o respingem H_O la nivelul de semnificație 0,05 este 1,96. Valoarea pe care am obținut-o este 3,74. Acesta este în mod clar un Z suficient de mare pentru a respinge H_O cel puțin la nivelul .01. Valoarea critică pentru ca Z să respingă H_O la nivelul 0,01 este 2,58.

Rețineți că am ajuns la aceeași concluzie folosind testul chi-pătrat din Secțiunea 3.1. De fapt, $Z^2 = X^2 = (3,74)^2 = 13,99$ și chi-pătratul necorectat pe care l-am calculat a fost 13,94 (diferența se datorează erorilor de rotunjire). Desigur, valorile critice ale lui X^2 și Z trebuie căutate în tabelele corespunzătoare. Unele valori apar în Anexa A.

3.18 t-Test pentru diferența dintre mijloacele a două grupuri independente: principii

Când am vrut să comparăm două grupuri pe o anumită măsură care a fost o variabilă discretă sau categorică, cum ar fi mortalitatea în două grupuri, am folosit testul chi-pătrat, descris în Secțiunea 3.1. Sau am putea folosi un test între proporții așa cum este descris în Secțiunea 3.17. Acum discutăm o metodă de comparare a două grupuri atunci când măsura interesului este o variabilă continuă.

Să luăm ca exemplu comparația vârstelor la prima sarcină a două grupuri de femei: cele care sunt avocate și cele care sunt asistente juridice. Un astfel de studiu ar putea fi de interes sociologic sau ar putea fi de interes pentru firmele de avocatură sau poate pentru o companie de alimente pentru copii care încearcă să-și concentreze strategia de publicitate mai eficient.

Presupunând că am luat eșantioane adecvate din fiecare grup, acum avem două seturi de valori: vârstele avocaților (grupa A) și vârstele parajuriștilor (grupa B), și avem o vârstă medie pentru fiecare eșantion. Am stabilit ipoteza noastră nulă după cum urmează:

H_O: „Vârsta medie a populației de avocați de la care noi am extras eșantionul A este același cu vârsta medie a populației de parajuriști din care am extras eșantionul B."

Ipoteza noastră alternativă este

H_A: „Vârstele medii ale celor două populații pe care le-am eșantionat sunt diferit."

În esență, *am extras eșantioane pe baza cărora vom face inferențe despre populațiile din care provin* . Suntem supuși acelorași tipuri de erori de tip I și de tip II pe care le-am discutat anterior.

Abordarea generală este următoarea. Știm că există o variabilitate a scorurilor în grupul A în jurul mediei pentru grupul A și în cadrul grupului B în jurul mediei pentru grupul B, pur și simplu pentru că chiar și într-o anumită populație, oamenii variază. Ceea ce dorim să aflăm este dacă variabilitatea dintre cele două medii ale eșantionului în jurul mediei mari a tuturor scorurilor este mai mare decât variabilitatea vârstelor din cadrul grupurilor în jurul propriilor medii. Dacă există la fel de multă variabilitate în cadrul grupurilor ca și între grupuri, atunci probabil că aceștia provin din aceeași populație.

Testul potrivit aici este testul t. Calculăm o valoare cunoscută sub numele de t, care este egală cu diferența dintre cele două medii de eșantion împărțită la o eroare standard corespunzătoare. Standardul adecvat

eroarea se numește eroarea standard a diferenței dintre două medii și se scrie ca

$$S_{e} \sim X_1 - X_2$$

Distribuția lui t a fost tabelată și din tabele putem obține probabilitatea de a obține o valoare a lui t la fel de mare ca cea pe care am obținut-o de fapt, în ipoteza că ipoteza noastră nulă (fără diferențe între medii) este adevărată. Dacă această probabilitate este mică (adică, dacă este puțin probabil ca doar din întâmplare să obținem o valoare a t atât de mare dacă ipoteza nulă ar fi adevărată) am respinge ipoteza nulă și am accepta ipoteza alternativă că există într-adevăr o diferență între mediile populațiilor din care am extras cele două eșantioane.

3.19 Cum se face un test t: un exemplu

Deși testele t pot fi efectuate cu ușurință pe computere personale, un exemplu de calcule și interpretare este dat mai jos. Acest test statistic este efectuat pentru a compara mediile a două grupuri în ipoteza că ambele eșantioane sunt aleatorii, independente și provin din populații distribuite normal cu variațe necunoscute, dar egale .

Ipoteza nulă: $m_A = m_B$, sau echivalentul: $m_A - m_B = 0$.

Ipoteza alternativă: $m_A \quad m_B$, sau echivalentul: $m_A - m_B \quad 0$.

[Notă: Atunci când ipoteza alternativă nu specifică direcția diferenței (prin afirmarea, de exemplu, că m_A este mai mare decât m_B), ci pur și simplu spune că diferența *nu este egală cu 0,* se numește test cu două cozi. Când este specificată direcția diferenței, se numește test cu o singură coadă. Mai multe despre acest subiect apar în Secțiunea 5.4.]

$$t = \frac{X_A - X_B}{S_{X_A - X_B}}$$

Vârstele eșantionului A			Vârstele eșantionului B		
x_i	$(X_I - X_A$	$(X_{i-x_A})^2$	X	$(X_{i-X_B}$	$(Xi_{-XB})^2$
28	-3	9	24	2.4	5,76
30	-1	1	25	3.4	11.56
27	-4	16	20	-1,6	2,56
32	1	1	18	-3,6	12.96
34	3	9	21	-0,6	0,36
36	5	25	S = 108	S = 0	S = 33,20
30	-1	1			
S = 217	S = 0	S = 62			

$Media_A - \bar{X}A \sim\sim\wedge\sim\sim 7\sim^{-31}$; $Media_B = X_B -$ $\frac{S X_i}{} - 21_6 \cdot 6$

(Indicele *i* se referă la scorul *i* și este o convenție folosită pentru a indica faptul că însumăm toate scorurile.)

Numătorul lui t este diferența dintre cele două medii:

$$31 - 21{,}6 = 9{,}4$$

Pentru a obține numitorul lui t trebuie să calculăm eroarea standard a diferenței dintre medii, ceea ce facem după cum urmează:

Mai întâi obținem estimarea cumulativă a abaterii standard. Noi calculăm:

$$s_p \quad \frac{\wedge (X_i - X_A)_2 + X_i - X_B)_2}{n_A + n_B - 2} \qquad 62 + 33.20 \quad 7 + 5 - 2$$

$$\frac{95.20}{10} \quad 79\check{A}2 - \quad 3.09$$

$$S_{\bar{x}_A - \bar{x}_B} \quad S_p \quad \frac{X + X = 3.09}{n_B} \; n_A \quad + I = 3.09y/.3428 = 3.09 \text{ x}.5854 = 1.81$$

$$t = \frac{\bar{x}_A - \bar{x}_B}{S_{\bar{x}_A - \bar{x}_B}} = \frac{9.4}{1.81} = 5.19$$

Acest t este semnificativ la nivelul .001, ceea ce înseamnă că ați obține o valoare a t la fel de mare ca aceasta sau mai mare doar o dată dintr-o mie din întâmplare dacă ipoteza nulă ar fi adevărată. Așa că respingem ipoteza nulă a lipsei de diferență, acceptăm ipoteza alternativă și concluzionăm că avocații sunt mai în vârstă la prima sarcină decât parajuriștii.

3.20 Testul t de perechi potrivite

Dacă aveți o situație în care scorurile dintr-un grup se corelează cu scorurile din celălalt grup, nu puteți utiliza testul t obișnuit, deoarece presupune că cele două grupuri sunt independente. Această situație apare atunci când iei două măsuri asupra aceluiași individ. De exemplu, să presupunem că grupa A reprezintă scorurile de citire ale unui grup de copii luate la momentul 1. Acești copii au primit apoi instruire specială în citire pe o perioadă de șase luni și rezultatele lor de citire sunt din nou măsurate pentru a vedea dacă au obținut vreun câștig la momentul 2. Într-o astfel de situație, ați folosi un test t perechi potrivit.

Copil	O Scorurile inițiale de citire ale copiilor	B Scoruri de aceiași copii după 6 luni de antrenament	d = B - A	$d - \bar{d}$	$(d - \bar{d})^2$
(1)	60	62	2	1.4	1,96
(2)	50	54	4	3.4	11.56
(3)	70	70	0	-0,6	0,36
(4)	80	78	-2	-2,6	6,76
(5)	75	74	-1	-1,6	2,56
		Suma	3	0	23,20
Diferența medie = $\bar{d}$ = 3/5 = 0,60					

Ipoteza nulă: diferența medie = 0.

Ipoteza alternativă: diferența medie este mai mare decât 0.

$$\frac{d}{Sd'} \qquad \frac{s}{\sqrt{n}}$$

$$\frac{2\,3\,.\,20}{4}$$

$$S_d \quad \begin{matrix} 2.41 & 2.41 \\ 75 & 2.23 \end{matrix} = 1.08$$

$$t = \frac{.\,60}{1\,.\,08} = .\,56$$

Acest t nu este semnificativ, ceea ce înseamnă că nu respingem ipoteza nulă și concluzionăm că diferența medie a scorurilor la citire ar putea fi zero; adică este posibil ca programul de lectură de șase luni să nu fie - eficient. (Sau poate că studiul nu a fost suficient de mare pentru a detecta o diferență și am comis o eroare de tip II.)

Când diferența reală dintre perechile potrivite nu este în sine un număr semnificativ, dar cercetătorul poate *clasifica* scorurile diferențelor (ca fiind mai mari sau mai mici pentru perechile date). Testul potrivit este testul Wilcoxon pentru sumele de rang pentru perechi. Acesta este cunoscut ca un *test nonparametric* și, împreună cu alte astfel de teste, este descris cu o claritate rafinată în cartea clasică a lui Sidney Siegel, *Statistics nonparametrice pentru științe comportamentale* (enumerate în Lecturile sugerate).

3.21 Când să nu faci o mulțime de teste t: Problema testelor multiple de semnificație

Un test t este utilizat pentru a compara mediile a două grupuri. Când există trei sau mai multe mijloace de grup de comparat, testul t nu este

adecvat. Pentru a înțelege de ce, trebuie să ne invocăm cunoștințele despre combinarea probabilităților din Secțiunea 2.2.

Să presupunem că testați efectele a trei tratamente diferite pentru hipertensiune arterială. Pacienții dintr-un grup A primesc un singur medicament, un diuretic; pacienții din grupa B primesc un alt medicament, un beta-blocant; iar pacientii din grupul C primesc o pastila placebo. Vrei să știi dacă oricare dintre medicamente este mai bun decât placebo în scăderea tensiunii arteriale și dacă cele două medicamente sunt diferite unul de celălalt în ceea ce privește efectul lor de scădere a tensiunii arteriale.

Există trei comparații care pot fi făcute: grupul A față de grupul C (pentru a vedea dacă diureticul este mai bun decât placebo), grupul B față de grupul C (pentru a vedea dacă beta-blocantul este mai bun decât placebo) și grupul A față de grupul B (pentru a vedea care dintre cele două medicamente active are mai mult efect). Ne-am stabilit nivelul de semnificație la .05, adică suntem dispuși să *greșim* respingând ipoteza nulă a lipsei de diferență între două medii, cu o probabilitate de .05 sau mai mică (adică, probabilitatea noastră de a face o eroare de tip I nu trebuie să fie mai mare de .05). Luați în considerare următoarele:

Comparație	Probabilitatea erorii de tip I	Probabilitatea de a *nu* face o eroare de tip I = 1 - *P* (eroare de tip I)
1. A vs. C	.05	1 - .05 = .95
2. B vs. C	.05	1 - .05 = .95
3. A vs. B	.05	1 - .05 = .95

Probabilitatea de *a nu* face o eroare de tip I la prima comparație *și* de a nu o face în a doua comparație *și* de a nu o face în a treia comparație = .95 x .95 x .95 = .86. Ne uităm aici la apariția *în comun* a trei evenimente (cele trei moduri de *a nu* comite o eroare de tip I) și combinăm aceste probabilități prin înmulțirea probabilităților individuale. (Nu uitați, când vedem „și” în contextul combinării probabilităților, înmulțim, când vedem „sau” adăugăm.) Așadar, acum, știm că probabilitatea globală de a *nu* comite o eroare de tip I în oricare dintre cele trei comparații posibile este de 0,86.

Prin urmare, probabilitatea de a comite o astfel de eroare este 1 - probabilitatea de a nu o comite, sau 1 - .86 = .14. Astfel, probabilitatea globală a unei erori de tip I ar fi considerabil mai mare decât .05 pe care am specificat-o ca fiind dezirabil. De fapt, numerele sunt puțin diferite deoarece cele trei comparații nu sunt evenimente independente, deoarece aceleași grupuri sunt utilizate în mai multe comparații, astfel încât combinarea probabilităților în această situație nu ar implica regula simplă a înmulțirii pentru apariția în comun a evenimentelor independente. Cu toate acestea, este suficient de aproape pentru a ilustra faptul că efectuarea de comparații multiple în același experiment are ca rezultat un nivel de semnificație destul de diferit (.14 în acest exemplu) decât cel pe care l-am ales (.05). Când există mai mult de trei grupuri de comparat, situația se înrăutățește.

3.22 Analiza variației: comparație între mai multe grupuri

Tehnica adecvată pentru analiza variabilelor continue atunci când există trei sau mai multe grupuri de comparat este analiza variației , denumită în mod obișnuit ANOVA. Un exemplu ar putea fi compararea efectelor de reducere a tensiunii arteriale ale celor trei medicamente.

3.23 Principii

Principiile implicate în analiza variației sunt aceleași cu cele din testul t. În ipoteza nulă, am avea următoarea situație: ar exista o singură populație mare și dacă am alege eșantioane de o anumită dimensiune din acea populație, am avea o grămadă de medii de eșantionare care ar varia în funcție de șansă în jurul mediei mari a întregii populații. Dacă se dovedește că variază în jurul valorii mari mai mult decât ne-am aștepta doar din întâmplare, atunci poate că altceva decât șansa funcționează. Poate că nu toți provin din aceeași populație. Poate că ceva distinge grupurile pe care le-am ales. Apoi am respinge ipoteza nulă conform căreia toate mijloacele sunt egale și am concluziona că mijloacele sunt diferite

 unii pe alții, mai mult decât întâmplător. În esență, dorim să știm dacă variabilitatea mediilor tuturor grupurilor este substanțial mai mare decât variabilitatea în cadrul fiecăruia dintre grupuri în jurul propriei medii.

Calculăm o cantitate cunoscută sub numele de *varianță între grupuri,* care este variabilitatea mediilor grupului în jurul mediei mari a tuturor datelor. Calculăm o altă cantitate numită *varianță în interiorul grupurilor* , care este variabilitatea scorurilor din cadrul fiecărui grup în jurul propriei sale medii. Una dintre ipotezele analizei variației este că amploarea variabilității indivizilor în cadrul grupurilor este aceeași pentru fiecare dintre grupuri, astfel încât putem pune în comun estimările variațiilor individuale în cadrul grupului pentru a obține o estimare mai fiabilă a variației globale în cadrul grupurilor. Dacă există atâta variabilitate a indivizilor în *cadrul* grupurilor cât există variabilitate a mediilor *între* grupuri, mediile provin probabil din aceeași populație, ceea ce ar fi în concordanță cu ipoteza lipsei unei diferențe adevărate între medii, adică nu am putea respinge ipoteza nulă a lipsei de diferență între medii.

Raportul dintre variația între grupuri și varianța în interiorul grupurilor este cunoscut sub numele de raportul F. Valorile distribuției F apar în tabele în multe texte statistice și dacă valoarea obținută din experimentul nostru este mai mare decât *valoarea critică* care este tabelată, atunci putem respinge ipoteza lipsei de diferență.

Există diferite valori critice ale lui F în funcție de câte grupuri sunt comparate și de câte scoruri există în fiecare grup. Pentru a citi tabelele lui F, trebuie să cunoaștem cele două valori ale gradelor de libertate (df). Df corespunzătoare variației între grupuri, care este numărătorul raportului F, este egal cu k - 1, unde k este numărul de grupuri. Df-ul corespunzător numitorului raportului F, care este varianța în interiorul grupurilor, este egal cu k x (n - 1), adică numărul de grupuri înmulțit cu numărul de scoruri din fiecare grup minus unu. De exemplu, dacă în experimentul nostru de hipertensiune există 100 de pacienți în fiecare dintre cele trei grupe de medicamente, atunci numărul de grade de libertate ar fi 3 - 1 = 2, iar numitorul gradelor de libertate ar fi 3 x 99 = 297. Un raport F ar trebui să fie de cel puțin 3,07 pentru un nivel de semnificație de 0,05. Dacă ar fi comparate patru grupuri, atunci gradele de libertate ale numărătorului ar fi 3,

iar valoarea critică a lui F ar trebui să fie 2,68. Dacă nu există un număr egal de indivizi în fiecare grup, atunci numitorul gradelor de libertate este (n_1 - 1) + (n_2 - 1) + (n_3 - 1).

Nu vom prezenta aici calculele efective necesare pentru a face un test F pentru ca in zilele noastre acestea se fac rareori manual. Există un număr mare de programe disponibile pentru computerele personale care pot efectua teste F, teste t și majoritatea altor analize statistice. Cu toate acestea, mai jos este prezentat tipul de rezultate la care se poate aștepta de la aceste programe . Sunt prezentate date rezumative din studiul TAIM (Trial of An tihypertensive Interventions and Management). Studiul TAIM a fost conceput pentru a evalua efectul dietei și al medicamentelor, utilizate singure sau în combinație între ele, pentru a trata persoanele supraponderale cu hipertensiune ușoară (tensiune arterială ridicată). [10,11]

Următorul tabel arată scăderea medie a tensiunii arteriale după șase luni de tratament cu fiecare medicament, numărul de persoane din fiecare grup și abaterea standard a *modificării* tensiunii arteriale în fiecare grup.

Grup de droguri	*n*	Scădere medie (în unități de tensiune arterială diastolică după 6 luni de tratament)	Abaterea standard
A. Diuretic	261	12.1	7.9
B. Beta-blocant	264	13.5	8.2
C. Placebo	257	9.8	8.3

Următorul tabel rezultă dintr-o analiză a varianței datelor din acest studiu. Acesta trebuie interpretat astfel:

ANOVA					
Sursa de variație	Grade de libertate	Suma pătratelor	Pătrat mediu	Raportul F	*P* - > F

Între grupuri	2	1776,5	888,2	13.42	.0001
În grupuri	779	5256,9	66.2		
	781				

Pătratul mediu este suma pătratelor împărțită la gradele de libertate. Pentru între grupuri, este variația mediei de grup în jurul mediei mari, în timp ce pentru grupurile în interior este estimarea cumulativă a variației scorurilor individuale în jurul valorii medii ale grupului respectiv. Pătratul mediu din interiorul grupurilor se mai numește și pătratul mediu al erorii. (Un punct important este că rădăcina pătrată a pătratului mediu al erorii este estimarea combinată a deviației standard în interiorul grupurilor . În acest caz, este $766 . 2 = 8 . 14$. Este aproximativ echivalentă cu abaterea standard medie.) F este raportul dintre și mediile pătratelor; în acest exemplu este 888,2/66,2 = 13,42.

Raportul F este semnificativ la nivelul .0001, deci putem respinge ipoteza nulă că *toate* mediile de grup sunt egale. Cu toate acestea, nu știm unde este diferența. Este grupul A diferit de grupul C, dar nu de grupul B? Nu ar trebui pur și simplu să facem posibile toate comparațiile în perechi din cauza problemei comparațiilor multiple discutate mai sus. Dar există modalități de a gestiona această problemă. Una dintre ele este procedura Bonferroni, descrisă în secțiunea următoare.

3.24 Procedura Bonferroni: O abordare pentru a face comparații multiple

Aceasta este o modalitate de a gestiona problema comparațiilor multiple. Procedura Bonferroni implică că, dacă de exemplu facem cinci comparații, probabilitatea ca *niciuna* dintre cele cinci valori *p* să nu scadă sub .05 este

de cel puțin 1 - (5 x .05) = .75 atunci când ipoteza nulă a mediei egale este cu adevărat adevărată. Aceasta înseamnă că există o probabilitate de până la .25 ca cel puțin o valoare *p* să atingă nivelul de semnificație de .05 doar întâmplător, *chiar dacă tratamentele nu diferă într-adevăr* . Pentru a ocoli acest lucru, împărțim nivelul de semnificație general ales la numărul de comparații bidirecționale care trebuie făcute, considerăm că această valoare este nivelul de semnificație pentru orice comparație unică și respingem ipoteza nulă a lipsei de diferență numai dacă atinge acest nou nivel de semnificație.

De exemplu, dacă dorim un nivel de semnificație general de .05 și vom face trei comparații între medii, ar trebui să atingem nivelul .05/3 = .017 pentru a respinge ipoteza nulă și a concluziona că există o diferență între cele două medii. Este necesară o bună autodisciplină pentru a respecta această procedură și pentru a nu declara că diferența dintre două mijloace este puțin probabil să se datoreze întâmplării dacă comparația particulară are semnificație la *p* = .03, să zicem, în loc de .017. Procedura Bonferroni nu necesită un test F prealabil. Să aplicăm procedura Bonferroni datelor noastre.

Mai întâi comparăm fiecare dintre medicamente cu placebo. Se calculează t pentru diferența dintre mediile grupului A față de grupul C.

$$t = \frac{\bar{x}_A - \bar{X}_C}{S_{\bar{X}_A - \bar{X}_C}}$$

$$s.e._{\bar{x}_A - \bar{x}_C} = s_p\sqrt{\frac{1}{n_A} \quad \frac{1}{n_C}}$$

$$\frac{12.1 - 9.8}{8.14\sqrt{\frac{1}{261} + \frac{1}{257}}} = \frac{2.3}{.715} = 3.22$$

$$p = .0014$$

Rețineți că folosim 8.14 ca *s* pooled. Am obținut acest lucru din analiza varianței ca estimare a abaterii standard comune. Gradele de libertate pentru a intra în tabelele t sunt 261 + 257 - 2 = 516.

Se pare că probabilitatea de a obține o valoare t atât de mare din întâmplare este de numai .0014. Putem spune cu siguranță că diureticul reduce tensiunea arterială mai mult decât placebo. Același lucru este valabil și pentru comparația dintre beta-blocant și placebo. Acum să comparăm cele două medicamente, B versus A:

$$t = \frac{13.5 - 12.1}{8.14\sqrt{\frac{1}{264} + \frac{1}{261}}} = \frac{1.4}{AH} = 1.97$$

Valoarea *p* corespunzătoare acestui t value is .049. It might be attentant de a declara o diferență semnificativă the .05 level, but remem-
după procedura Bonferroni necesită să
obținem o valoare *p* de .017 sau mai puțin pentru semnificație ajustată pentru comparații multiple. Valoarea critică a lui t corespunzătoare lui *p* = .017 este 2,39 și am ajuns doar la 1,97. Recent, [12] s-au pus sub semnul întrebării ajustările de rutină pentru comparații multiple, pe motiv că, prin urmare, putem comite mai multe erori de tip II și pierdem efecte importante. În orice caz, nivelurile *p* ar trebui raportate, astfel încât cititorul informat să poată evalua dovezile .

3.25 Analiza variației atunci când există două variabile independente: ANOVA cu doi factori

Exemplul de mai sus este denumit ANOVA unidirecțională deoarece puteți împărți toate scorurile într-un singur fel, în funcție de grupul de medicamente căruia i-au fost alocați pacienții. Grupul de droguri se numește „factor" și acest factor are trei niveluri, adică există trei categorii de droguri. Totuși, poate exista un alt factor care clasifică indivizii și, în acest caz, am avea un ANOVA cu două sensuri sau cu doi factori. În experimentul pe care l-am folosit ca exemplu, pacienții au fost repartizați la unul dintre cele trei medicamente menționate mai sus, precum și la unul dintre cele trei regimuri de dietă - scădere în greutate, restricție de sodiu (sare) sau nicio schimbare față de dieta lor obișnuită, care este analogă cu o condiție de dietă placebo. Diagrama de mai jos ilustrează acest design cu doi factori și scăderea medie a tensiunii arteriale în fiecare grup, precum și numărul de cazuri din fiecare

grup, care sunt prezentate în paranteză.

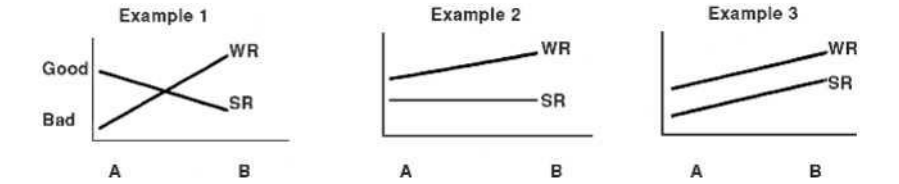
	Dieta			
Medicament	De obicei	Reducere în greutate	Restricție de sodiu	Total
Diuretic	10.2 (87)	14.5 (86)	11.6 (88)	12.1 (261)
Beta-blocant	12.8 (86)	15.2 (88)	12.6 (90)	13.5 (264)
Placebo	8.7 (89)	10.8 (89)	10.1 (79)	9.8 (257)
	Total	10,5 (262)	13,5 (263)	11,5 (257)

Acum suntem interesați să comparăm cele trei mijloace care reprezintă modificarea tensiunii arteriale în grupurile de medicamente, cele trei mijloace care reprezintă schimbări în grupurile de dietă și interacțiunea dintre medicament și dietă. Acum explicăm conceptul de interacțiune.

3.26 Interacțiunea între două variabile independente

Interacțiunea dintre două variabile independente se referă la diferențele în efectul unei variabile în funcție de nivelul celei de-a doua variabile. De exemplu, poate un medicament produce efecte mai bune atunci când este combinat cu o dietă de scădere în greutate decât atunci când este combinat cu o dietă cu conținut limitat de sodiu. Este posibil să nu existe un efect semnificativ al acelui medicament atunci când toate grupurile de dietă sunt grupate, dar dacă ne uităm la efectele separat pentru fiecare grup de dietă, putem descoperi o interacțiune între cei doi factori: dietă și medicament.

Diagramele de mai jos ilustrează conceptul de efecte de interacțiune. WR înseamnă scădere în greutate, iar SR înseamnă restricție de sodiu (sare).



În exemplul 1, medicamentul B este mai bun decât medicamentul A la cei cu scădere în greutate, dar la cei cu restricție de sare, medicamentul A este mai bun decât medicamentul B. Dacă am compara doar media pentru medicamentul A, combinând diete, cu media pentru medicamentul B, ar trebui să spunem că nu există nicio diferență între medicamentul A și medicamentul B, dar dacă ne uităm la cele două diete separat, vedem efecte destul de diferite ale celor două medicamente.

În exemplul 2, nu există nicio diferență între cele două medicamente pentru cei care restricționează sare, dar există un efect mai mic al medicamentului A decât al medicamentului B pentru cei care slăbesc.

În exemplul 3, nu există nicio interacțiune; există un efect egal pentru ambele diete: cele două linii sunt paralele; pantele lor sunt aceleași. Medicamentul B este mai bun decât medicamentul A atât pentru cei care slăbesc, cât și pentru cei care se limitează la sare.

3.27 Exemplu de ANOVA bidirecțional

În continuare este un tabel de date din studiul TAIM care arată rezultatele unei *analize bidirecționale a varianței:*

ANOVA bidirecțional din datele TAIM

Sursă	DF	Suma ANOVA a pătratelor	Pătrat mediu	Valoarea F	Probabilitate
Grup de droguri	2	1776,49	888,25	13.68	.0001
Grup de dieta	2	1165,93	582,96	8,98	.0001
Medicament x	4	214,50	53,63	0,83	.509

Eroare	773	50.185,46	64,93		

Rețineți că pătratul mediu al erorii aici este 64,93 în loc de 66,9 când am făcut analiza unidirecțională. Acest lucru se datorează faptului că am explicat o parte din variația erorii ca fiind datorată efectelor dietei și efectelor de interacțiune (am „eliminat" aceste efecte din variația erorii). Astfel, 64,93 reprezintă varianța datorată erorii pure sau varianța inexplicabilă. Acum putem folosi rădăcina pătrată a acesteia, care este 8,06, ca estimare a abaterii standard comune. Explicăm rezultatele după cum urmează: Există un efect semnificativ al medicamentului (p = 0,0001) și un efect semnificativ al dietei (p = 0,0001), dar nicio interacțiune a medicamentului cu dieta (p = 0,509).

Am făcut deja cele trei comparații pe perechi, prin teste t pentru diferența dintre două medii între medicamente (adică, placebo vs. diuretic, placebo vs. beta-blocant și diuretic vs. beta-blocant). Putem face același lucru pentru cele trei diete. Valorile lor medii sunt afișate mai jos:

Grup de dieta	***n***	Scădere medie a tensiunii arteriale diastolice	Abaterea standard
Reducere în greutate	263	13.5	8.3
Restricție de sodiu	257	11.5	8.3
Dieta obisnuita	262	10.5	8.0
(Estimarea comună a sd = 8,06)			

Dacă am face teste t, am constata că reducerea în greutate este mai bună decât dieta obișnuită (p = .0000), dar restricția de sodiu nu arată nicio - îmbunătățire semnificativă față de dieta obișnuită (p = .16).

Scăderea în greutate în comparație cu restricția de sodiu este, de asemenea, semnificativ mai bună cu p = .005, care este mult sub p = .017 cerut de procedura Bonferroni. (T pentru această comparație în perechi este 2,83, care este peste valoarea critică de 2,39.)

3.28 Testul Kruskal-Wallis pentru a compara mai multe grupuri

Analiza varianței este valabilă atunci când variabila de interes este continuă, provine dintr-o distribuție normală, adică curba familiară în formă de

clopot, iar variațiile în cadrul fiecăruia dintre grupurile comparate sunt în esență egale. De multe ori, totuși, trebuie să ne confruntăm cu situații în care dorim să comparăm mai multe grupuri pe o variabilă care nu îndeplinește toate condițiile de mai sus. Acesta ar putea fi un caz în care putem spune că o persoană este mai bună decât alta, dar nu putem spune exact cât de mult mai bună. Într-un astfel de caz, am clasifica oamenii și am compara grupurile utilizând testul Kruskal-Wallis pentru a determina dacă este probabil ca toate grupurile să provină dintr-o populație comună. Acest test este analog cu analiza unidirecțională a variației, dar în loc să folosească scorurile originale, folosește *clasamentele* scorurilor. Se numește *test neparametric* . Acest test este disponibil în multe programe de calculator, dar un exemplu apare în Anexa C.

3.29 Asociere și cauza: coeficientul de corelație

O clasă comună de probleme în acumularea și evaluarea dovezilor științifice este evaluarea asocierii a două variabile. Există o asociere între sărăcie și dependența de droguri? Stresul emoțional este asociat cu bolile cardiovasculare?

Pentru a determina asocierea, trebuie mai întâi să cuantificăm ambele variabile. De exemplu, stresul emoțional poate fi cuantificat prin utilizarea unui test psihologic adecvat al stresului sau prin definirea, evaluarea și evaluarea clară pe o scară a factorului de stres în situația de viață a unui individ, în timp ce hipertensiunea (definită ca o citire a tensiunii arteriale) poate fi considerată ca aspectul particular al bolii cardiovasculare care trebuie studiat. Când variabilele au fost cuantificate, trebuie calculată o măsură a asocierii pentru a determina puterea relației. Una dintre cele mai comune măsuri de asociere este *coeficientul de corelație* , *r* , care este un număr derivat din date care poate varia între -1 și +1. (Metoda de calcul apare în Anexa D.) Când $r = 0$ înseamnă că nu există asociere între cele două variabile. Un exemplu în acest sens ar putea fi corelația dintre tensiunea arterială și numărul de fire de păr de pe cap. Când $r = +1$, o corelație pozitivă perfectă , înseamnă că există o relație directă între cele două variabile: un individ care are un scor mare la o variabilă are și un scor mare la cealaltă, iar scorul unei variabile poate fi prezis exact din scorul celeilalte variabile. Acest tip de corelație există doar în modelele

deterministe, unde există într-adevăr o relație funcțională. Un exemplu ar putea fi corelația dintre vârsta unui copac și numărul de inele pe care le are. Un coeficient de corelație de -1 indică o relație inversă perfectă, unde un scor mare la o variabilă înseamnă un scor scăzut la cealaltă și unde, ca în corelația pozitivă perfectă , nu există nicio eroare de măsurare. Coeficienții de corelație între 0 și +1 și între 0 și -1 indică diferite forțe ale asociațiilor.

Acești coeficienți de corelație se aplică atunci când relația de bază dintre cele două variabile este liniară. Luați în considerare un grup de persoane pentru fiecare dintre care avem o măsurare a greutății față de înălțime; vom constata că putem trage o linie dreaptă prin puncte. Există o asociere liniară între greutate și înălțime și coeficientul de corelație ar fi pozitiv, dar mai mic de 1.

Diagramele din Figura 3.9 ilustrează reprezentări ale diferiților coeficienți de corelație.

3.30 Cât de sus este mare?

Răspunsul la această întrebare depinde de domeniul de aplicare, precum și de mulți alți factori. Dintre variabilele psihologice, care sunt greu de măsurat cu precizie și sunt afectate de multe alte variabile , corelațiile sunt în general (deși nu neapărat) mai mici decât în rândul variabilelor biologice unde este posibilă măsurarea mai precisă. Următorul exemplu vă poate oferi o idee despre ordinele de mărime. Corelațiile dintre aptitudinea verbală și aptitudinea nonverbală, măsurate pentru școlari din Philadelphia prin teste naționale standardizate, variază de la .44 la .71, în funcție de rasă și clasa socială a grupurilor. [13]

3.31 Căi cauzale

Dacă obținem o corelație semnificativă, atunci ne întrebăm ce situații ar putea fi responsabile pentru aceasta? Figura 3.10 ilustrează unele posibile structuri

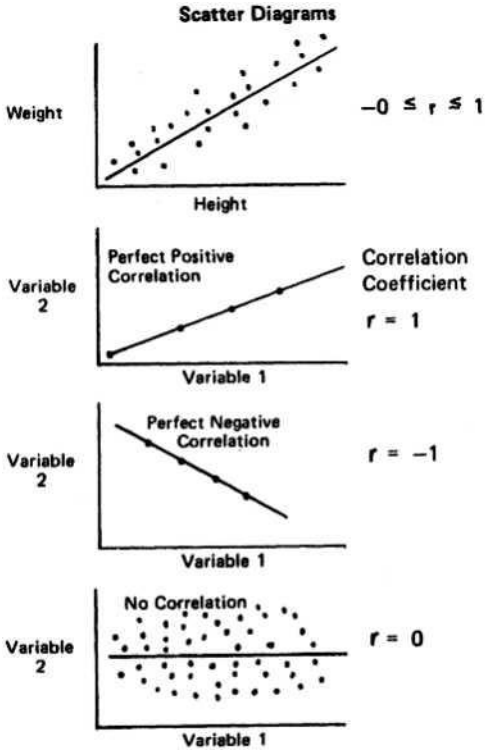


Figura 3.9

relații culturale care pot sta la baza unui coeficient de corelație semnificativ , așa cum sugerează Sokal și Rohlf. [14] Luăm în considerare două variabile, W (creșterea în greutate) și B (tensiunea arterială), iar r_{WB} reprezintă corelația dintre ele. Rețineți că numai în diagramele (1), (2) și (6) corelația dintre W și B apare din cauza unei relații cauzale între cele două variabile. În diagrama (1), W determină în întregime B; în diagrama (2), W este o cauză parțială a lui B; în diagrama (6), W este unul dintre câțiva - determinanți a lui B. În toate celelalte relații structurale, corelația dintre W și B apare din cauza influențelor comune asupra ambelor variabile. Astfel, trebuie subliniat că *existența unei corelații între două variabile nu implică neapărat cauzalitate.* corela-

pot apărea deoarece o variabilă este cauza parțială a alteia sau cele două variabile corelate au o cauză comună. Alți factori, cum ar fi eșantionarea, variația celor două populații și așa mai departe, afectează și dimensiunea coeficientului de corelație. Prin urmare, trebuie avută grijă la interpretarea acestor coeficienți.

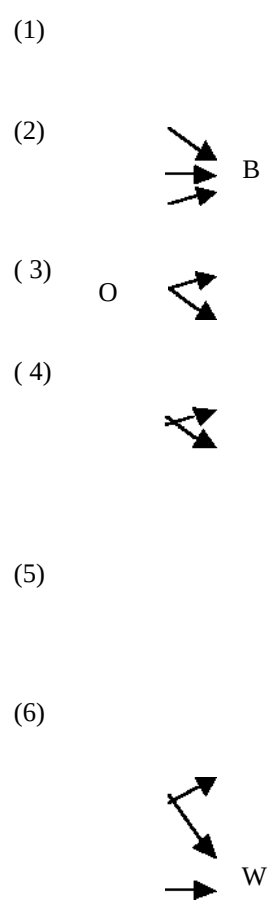

(1)
W = crestere in greutate; B = tensiunea arterială
W determină în întregime B
$r_{WB} = 1$
(2)
B
W este unul dintre câțiva determinanți ai lui B
A = vârsta; K = funcția rinichilor Aceste
variabile ar putea afecta și tensiunea arterială
r_{WB} este mai mic de 1
(3)
O
Cauza comună, vârsta, determină în totalitate
atât tensiunea arterială, cât și creșterea în
greutate $r_{WB} = 1$
(4)
Corelația dintre W și B se datorează unei cauze
comune, A (vârstă), dar C (aportul caloric) și K
(funcția rinichilor) determină și W (creșterea în
greutate) și B (respectiv tensiunea arterială).
r_{WB} este mai mic de 1
(5)
Corelația dintre W și B se datorează a două
cauze comune, A (vârsta) și factorii genetici G
r_{WB} este mai mic de 1
(6)
Corelația dintre W și B se datorează efectului
direct al W (creșterea în greutate) asupra B
(tensiunii arteriale) precum și unei cauze
comune, A (vârsta) care afectează ambele
variabile; G (factorii genetici) afectează și B (-
tensiunea arterială) și C (aportul caloric)
afectează W (creșterea în greutate) r_{WB} este mai
mic de 1
W

Figure 3.10

3.32 Regresia

Rețineți că în Figura 3.9 am trasat linii care par să se potrivească cel mai bine cu punctele de date. Acestea se numesc *linii de regresie.* Au următoarea formă:
$Y = a + bX$. În diagrama de dispersie de sus etichetată (a), *Y* este variabila dependentă greutatea și *X* , sau înălțimea, este variabila independentă. Spunem că greutatea este o funcție de înălțime. Mărimea *a* este interceptarea. Este locul în care linia traversează axa *Y.* Mărimea *b* este panta și este rata de modificare a lui *Y* pentru o modificare unitară a lui *X.* Dacă panta este 0, înseamnă că avem o dreaptă paralelă cu axa *x* , ca în ilustrația (d). De asemenea, înseamnă că nu putem prezice *Y* dintr-o cunoaștere a lui *X,* deoarece nu există nicio relație între *Y* și *X.* Dacă avem situația prezentată în diagramele de dispersie (b) sau (c), știm exact cum se schimbă *Y atunci când X* se schimbă și putem prezice perfect *Y* din cunoașterea lui *X* fără eroare. În diagrama de dispersie (a), putem vedea că pe măsură ce *X* crește, *Y* crește, dar nu putem prezice *Y* perfect pentru că punctele sunt împrăștiate în jurul liniei pe care am trasat-o. Putem, totuși, să punem limite de încredere în jurul predicției noastre, dar mai întâi trebuie să stabilim forma liniei pe care ar trebui să o trasăm prin puncte. Trebuie să estimăm valorile pentru interceptarea și panta. Acest lucru se face prin găsirea „liniei cele mai potrivite".

Linia care se potrivește cel mai bine punctelor are următoarele caracteristici: dacă luăm fiecare dintre punctele de date și calculăm distanța sa verticală față de linie și apoi pătram acea distanță, suma acelor distanțe pătrate va fi mai mică decât suma acestor distanțe pătrate de la orice altă linie pe care am putea-o trasa. Aceasta se numește potrivirea *celor mai mici pătrate* . Luați în considerare datele de mai jos, unde *Y* ar putea fi un scor la un test și *X* ar putea fi un scor la un alt test.

	Scor	
Individual	*X*	*Y*
O	5	7
B	8	4
C	15	8
D	20	10
E	25	14

Calculele pentru determinarea liniei celei mai potrivite sunt prezentate în Anexa D. Cu toate acestea, majoritatea pachetelor de calculatoare statistice pentru calculatoare personale oferă un program de regresie liniară care efectuează aceste calcule . Figura 3.11 ilustrează aceste puncte reprezentate într-un scattergram și arată linia celor mai mici pătrate.

Ecuația dreptei este $Y = 2{,}76 + .40\,X$. Intercepta a este 2,76, astfel încât linia traversează axa y la $Y = 2{,}76$. Panta este de .40. De exemplu, putem calcula un Y *prezis* pentru $X = 10$ pentru a obține

$$Y = 2{,}76 + (.40)(10) = 2{,}76 + 4 = 6{,}76$$

D *t ‗ urile* sunt distanțe de la puncte la linie. Suma acestor distanțe pătrate este mai mică pentru această linie decât ar fi pentru orice altă linie pe care am putea-o trasa.

Coeficientul de corelație pentru aceste date este .89. Pătratul coeficientului de corelație, r^2, poate fi interpretat ca proporția variației în Y care este explicată prin X . În exemplul nostru, $.89^2 = .79$; astfel

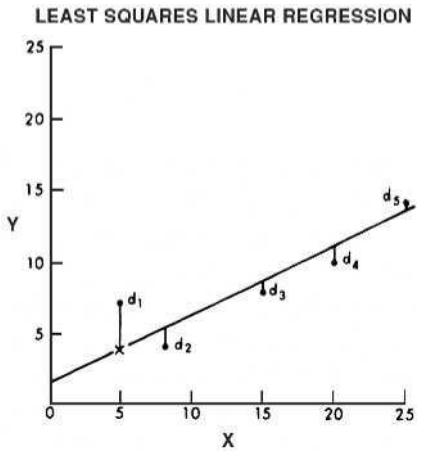


Figura 3.11

79% din variația lui *Y* este explicabilă prin variabila *X,* iar 21% nu este contabilizată.

3.33 Legătura dintre regresia liniară și coeficientul de corelație

Coeficientul de corelație și panta dreptei de regresie liniară sunt legate prin formule

$$r = b\frac{S_x}{S_y}, \qquad b = r\frac{S_y}{S_x}$$

unde s_x este abaterea standard a variabilei X , s_y este abaterea standard a variabilei Y , b este panta dreptei și r este coeficientul de corelație.

3.34 Regresia liniară multiplă

Când avem două sau mai multe variabile independente și o variabilă dependentă continuă, putem folosi analiza regresiei multiple. Forma pe care o ia aceasta este

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + bkXk$$

De exemplu, *Y* poate fi tensiunea arterială și X_1 poate fi vârsta, X_2 poate fi greutatea, X_3 poate fi antecedente familiale de hipertensiune arterială. Putem avea cât mai multe variabile, unde ultima variabilă este a k -a variabilă. B_i - *urile* sunt *coeficienți de regresie.* Rețineți că antecedentele familiale de hipertensiune arterială nu sunt o variabilă continuă. Poate fi fie da, fie nu. Numim aceasta variabilă dihotomică și o putem folosi ca orice altă variabilă într-o ecuație de regresie, atribuind un număr fiecăruia dintre cele două răspunsuri posibile, de obicei făcând un răspuns da = 1 și un răspuns nu = 0. Programele de calculator statistice includ de obicei regresia liniară multiplă.

Urmează un exemplu din studiul TAIM și este menit doar să vă ofere o idee despre cum să interpretați o ecuație de regresie multiplă. Această analiză se referă la grupul de 89 de persoane care au fost alocate unui - medicament pla cebo și unui regim de scădere în greutate. Variabila dependentă este modificarea tensiunii arteriale.

Variabilele independente sunt prezentate mai jos:

Variabilă	Coeficient: b_i	p
Intercepta	-15.49	.0016
Vârstă	.077	.359
Cursa 1 = Negru 0 = nenegru	4.22	.021
S ex 1 = Masculin 0 = Femeie	1,50	.390
Lire sterline pierdute	.13	.003

Notă: Sexul este codificat ca 1 dacă este bărbat și 0 dacă este femeie; rasa este codificată ca 1 dacă este neagră și 0 dacă nu este neagră; *p* este folosit pentru a testa dacă coeficientul este semnificativ diferit de 0. Ecuația, atunci, este

modificarea tensiunii arteriale =

—15,49 + 0,077 (vârstă) + 4,22 (rasă) + 1,5 (sex) + 0,13 (schimbarea greutății)

Vârsta nu este semnificativă (p = .359), nici sexul (p = .390). Cu toate acestea, rasa este semnificativă (p = .021), indicând faptul că negrii au avut o probabilitate mai mare decât cei nenegri de a avea o scădere a tensiunii arteriale, controlând simultan toate celelalte variabile din ecuație. Kilogramele pierdute sunt, de asemenea, semnificative, indicând că cu cât pierderea în greutate era mai mare, cu atât scăderea tensiunii arteriale era

mai mare.

3.35 Rezumat Până acum

Investigarea unei probleme științifice necesită adesea analize statistice, mai ales acolo unde există variabilitate în ceea ce privește caracteristicile de interes. Variabilitatea poate apărea din două surse: caracteristica poate fi variabilă în mod inerent în populație și/sau poate exista o eroare de măsurare.

În acest capitol am subliniat că pentru a evalua un program sau un medicament, pentru a compara două grupuri pe o anumită caracteristică, pentru a efectua o investigație științifică a oricărei probleme, este necesară cuantificarea variabilelor.

Variabilele pot fi cuantificate ca fiind discrete sau continue și există tehnici statistice adecvate care se ocupă de fiecare dintre acestea. Am luat în considerare aici testul chi-pătrat, intervalele de încredere, testul Z, testul t, analiza variației, corelația și regresia. Am subliniat că în testarea ipotezelor suntem supuși la două tipuri de erori : eroarea de a respinge o ipoteză atunci când de fapt aceasta este adevărată și eroarea de a accepta o ipoteză atunci când de fapt aceasta este falsă. Scopul unui studiu bine conceput este de a minimiza probabilitatea de a face aceste tipuri de erori. Statisticile nu vor înlocui un design experimental bun, dar este un instrument necesar pentru a evalua dovezile științifice obținute din studii bine concepute.

Filosofic vorbind, statistica este o reflectare a vieții în două aspecte importante : (1) ca și în viață, nu putem fi niciodată siguri de nimic (dar în statistică putem pune o cifră de probabilitate care descrie gradul de incertitudine a noastră) și (2) totul este un compromis - în statistică, între certitudine și precizie, sau între două tipuri de erori; în viață, ei bine, completează-ți propriile compromisuri.

Capitolul 4
MAI MULT DESPRE EPIDEMIOLOGIE

Medicina pentru a produce sănătate trebuie să examineze boala; și muzica pentru a crea armonie, trebuie să investigheze discordia.

Plutarh 46-

4.1 Utilizările epidemiologiei

Epidemiologia poate fi definită ca studiul distribuției sănătății și bolilor în *grupuri de oameni* și studiul factorilor care influențează această distribuție. Epidemiologia modernă cuprinde, de asemenea, evaluarea modalităților de diagnostic și terapeutice și furnizarea de servicii de îngrijire a sănătății. Există o progresie în procesul științific (de-a lungul dimensiunii creșterii credibilității dovezilor), de la observarea întâmplătoare, la formarea de ipoteze, la observarea controlată, la studii experimentale. Figura 4.1 este o reprezentare schematică a utilizărilor epidemiologiei. Instrumentele folosite în acest demers sunt în provincia epidemiologiei și biostatisticii. Tehnicile utilizate în aceste discipline permit „detectivilor medicali" atât să descopere o problemă medicală, să evalueze dovezile despre cauzalitatea sau etiologia acesteia, cât și să evalueze intervențiile terapeutice pentru combaterea problemei.

Epidemiologia descriptivă oferă informații despre tiparul bolilor, despre „cine are ce și cât de mult din el", informații care sunt esențiale pentru planificarea asistenței medicale și alocarea rațională a resurselor. Astfel de informații pot descoperi adesea modele de apariție care sugerează relații etiologice și pot duce la strategii preventive. Astfel de studii sunt de obicei de tip transversal și conduc la formarea de ipoteze care pot fi apoi testate în studii caz-control, prospective și experimentale. Studiile clinice și alte tipuri de studii controlate servesc la evaluarea modalităților terapeutice și a altor mijloace de intervenție și astfel, în cele din urmă, determină standardele de practică medicală, care la rândul lor au impact asupra deciziilor de planificare a asistenței medicale. În secțiunea următoare, vom lua în considerare concepte epidemiologice selectate.

UTILIZĂRI ALE EPIDEMIOLOGIEI

DESCRIPTIV: Distribuția SĂNĂTĂȚII
boli în CARE
populație ----------► PLANIFICARE
----------► subgrupuri
(Studii transversale)

APARIȚIE

DE MODELE
----------(sugerând) ----------1

ETIOLOGIC: FORMARE Asociaţii
printre —► OF
---------- ► variabile, IPOTEZE relaţii temporale
(caz-control şi studii prospective)
(sugerând)

EXPERIMENTAL: Testarea STANDARDELOR DE
interventii MEDICALE _____
► ► PRACTICĂ
(studii clinice)

Figura 4.1

4.2 Câteva concepte epidemiologice: rate de mortalitate

În 1900, cele trei cauze majore de deces erau gripa sau pneumonia , tuberculoza și gastroenterita. Astăzi, cele trei cauze majore de deces sunt bolile de inimă, cancerul și accidentele; al patrulea este accidente vasculare cerebrale. Decesele cauzate de accident vascular cerebral au scăzut dramatic în ultimele decenii, probabil datorită controlului îmbunătățit al hipertensiunii arteriale, una dintre principalele cauze.
mari factori de risc pentru accident vascular cerebral. Aceste modele în schimbare de mortalitate reflectă condițiile de mediu în schimbare, o trecere de la bolile acute la cele cronice și o populație îmbătrânită supusă bolilor degenerative. Știm acest lucru dintr-o analiză a *ratelor*.

Compararea ratelor definite între diferite subgrupuri de indivizi poate oferi indicii cu privire la existența unei probleme de sănătate și poate duce la specificarea condițiilor în care este probabil să apară și să înflorească această problemă de sănătate identificată.

La utilizarea tarifelor, trebuie reținute următoarele puncte:

(1) O rată este o proporție care implică un numărător și un numitor .
(2) Atât numărătorul, cât și numitorul trebuie să fie clar definite, astfel

încât să știți la ce grup (numitor) se referă rata dumneavoastră.

(3) Numătorul este conținut în (este o submulțime a) numitorului . Aceasta este în contrast cu un raport în care numărătorul se referă la un grup diferit de numitor.

Ratele de mortalitate se referă la numărul de decese care au loc într-un anumit subgrup de populație și oferă adesea unul dintre primele indicii ale unei probleme de sănătate. Următoarele definiții sunt necesare pentru a continua discuția:

Rata brută anuală a mortalității (sau rata mortalității) este:

$$\frac{\text{numărul } \textit{total} \text{ de decese pe parcursul unui an în populația expusă riscului}}{\text{populația expusă riscului (de obicei luată ca populație la mijlocul anului)}}$$

Rata de mortalitate anuală specifică cauzei este:

$$\frac{\text{numărul de decese survenite din cauza unei } \textit{anumite cauze} \text{ în cursul anului în populația expusă riscului}}{\text{populația expusă riscului (toți cei în viață la jumătatea anului)}}$$

Rata de mortalitate anuală specifică vârstei este:

$$\frac{\text{numărul de decese survenite la grupa de vârstă dată în cursul anului în populația cu risc}}{\text{populația expusă riscului (cei din acea grupă de vârstă trăiesc la mijlocul anului)}}$$

Un motiv pentru a lua populația la mijlocul anului ca numitor este că o populație poate crește sau scădea în timpul anului în cauză, iar populația de la jumătatea anului este o estimare a numărului mediu pe parcursul anului. Totuși, se poate vorbi de rate de mortalitate pe o perioadă de cinci ani, mai degrabă decât pe o perioadă de un an, și se poate defini populația expusă riscului ca fiind toți cei în viață la începutul perioadei.

4.3 Tarife ajustate în funcție de vârstă

Atunci când se compară ratele de mortalitate între două populații, trebuie să se țină seama de compoziția de vârstă a populațiilor. Deoarece persoanele în vârstă au un număr mai mare de decese la 1.000 de persoane, dacă o populație este puternic ponderată de persoanele în vârstă, rata brută a mortalității ar fi mai mare decât la o populație mai tânără și o comparație între cele două grupuri ar putea reflecta doar discrepanța de vârstă, mai degrabă decât o diferență intrinsecă a experienței mortalității. O modalitate de a rezolva această problemă este de a compara ratele de deces specifice vârstei, ratele de deces specifice unui anumit grup de vârstă. O altă modalitate care este utilă atunci când este necesară o cifră rezumată generală este utilizarea ratelor *ajustate în funcție de vârstă* . Acestea sunt rate ajustate la ceea ce *ar fi* dacă cele două populații comparate ar avea aceleași distribuții de vârstă ca o populație standard selectată în mod arbitrar.

De exemplu, tabelul de mai jos arată ratele brute de mortalitate și ajustate în funcție de vârstă pentru Statele Unite ale Americii în cinci perioade de timp.15,7 Ajustarea se face la distribuția pe vârstă a populației în 1940, precum și la distribuția pe vârstă a populației în 2000. Astfel, vedem că în 1991 rata ajustată în funcție de vârstă a fost 5,0091/1 , dar ajustată la 0,0091/1. rata brută de mortalitate a fost de 8,6/1000. Aceasta înseamnă că dacă în

1991 distribuția pe vârstă a populației era aceeași ca în 1940, atunci rata mortalității ar fi fost de doar 5,1/1000 persoane. Ratele brute și ajustate în funcție de vârstă pentru 1940 sunt aceleași, deoarece populația din 1940 servește ca populație „standard" a cărei distribuție de vârstă este folosită ca bază pentru ajustare.

Când a fost ajustată la standardul anului 2000, rata ajustată în funcție de vârstă a fost de 9,3. Dacă în 1991 distribuția pe vârstă ar fi aceeași ca în 2000, atunci rata mortalității ar fi fost de 9,3/1000 de persoane. Deci ratele ajustate în funcție de vârstă depind de populația standard utilizată pentru ajustare. Rețineți că rata ajustată în funcție de vârstă pe baza populației în anul 2000 este mai mare decât rata ajustată în funcție de vârstă pe baza populației din 1940; acest lucru se datorează faptului că populația este mai în vârstă în anul 2000.

An	Rata brută de mortalitate la 1.000 de persoane	Rata ajustată în funcție de vârstă (la Populația în 1940)	Rata ajustată în funcție de vârstă (la populație în 2000)
1940	10.8	10.8	17.9
1960	9.5	7.6	13.4
1980	8.8	5.9	10.4
1991	8.6	5.1	9.3
2001	8.5	necalculat după 1998	8.6

Deși atât ratele brute, cât și cele ajustate în funcție de vârstă au scăzut din 1940, scăderea ratei ajustate în funcție de vârstă este mult mai mare. Modificarea procentuală a mortalității brute între 1940 și 1991 a fost (10,8 - 8,6)/10,8 = 20,4%, în timp ce modificarea procentuală a ratei ajustate în funcție de vârstă a fost (10,8 - 5,1)/10,8 = 0,528 sau 52,8%.

Motivul pentru aceasta este că populația este în creștere. De exemplu, proporția persoanelor de 65 de ani și peste s-a dublat între 1920 și 1960, crescând de la 4,8% din populație în 1920 la 9,6% în 1969. După 1998, Centrul Național de Statistică în Sănătate a folosit populația din 2000 ca populație standard față de care s-au făcut ajustări. Rata brută și rata mortalității ajustată în funcție de vârstă în anul 2001 sunt similare, și asta pentru că distribuția de vârstă în

Anul 2001 este similar cu distribuția pe vârstă din 2000, așa că ajustarea vârstei nu schimbă cu adevărat rata mortalității.

Ratele ajustate în funcție de vârstă sunt numere fictive - nu vă spun câți oameni au murit efectiv la 1.000, ci câți *ar fi* murit dacă componența vârstei ar fi fost aceeași în cele două populații. Cu toate acestea, ele sunt adecvate în scopuri de comparație. Metodele de efectuare a ajustării în funcție de vârstă sunt descrise în Anexa E.

4.4 Ratele de incidență și prevalență

Rata de prevalență și rata de incidență sunt două măsuri ale morții (boală).

Rata de prevalență a unei boli este definită ca

$$\frac{\text{Numărul de persoane cu o boală}}{\text{Numărul total de persoane din populație expuse riscului la un anumit moment în timp}}$$

(Acest lucru este cunoscut și sub denumirea de *prevalență punctuală* , dar mai general denumită doar „prevalență.") De exemplu, prevalența hipertensiunii arteriale în 1973 în rândul bărbaților de culoare, cu vârste cuprinse între 30 și 69 de ani, definită ca o tensiune arterială diastolică - (DBP) de 95 mm Hg sau mai mult, la un program de screening a tensiunii arteriale condus de Programul de detectare a hipertensiunii arteriale și de monitorizare a hipertensiunii arteriale (^{FPU}) fii:

$$\frac{4,268 \text{ } \textit{cu DBP > 95 mmHg}}{15,190 \text{ } \textit{de bărbați de culoare cu vârsta cuprinsă între } 30 \text{ și } 69 \text{ de ani } \textit{au fost examinați}} \times 100 = 28,1 \text{ la } 100$$

Mai multe puncte trebuie remarcate cu privire la această definiție:

(1) Grupul de risc (numitorul) este definit clar ca bărbați de culoare, cu vârste cuprinse între 30-69 de ani.

(2) Momentul de timp este specificat ca timp de screening.

(3) Definiția bolii este clar specificată ca o tensiune arterială diastolică de 95 mm Hg sau mai mare. (Acesta poate include persoanele care sunt tratate pentru boală, dar a căror presiune este încă mare și cei care nu sunt tratați.)
(4) Numărătorul este subgrupul de indivizi din grupul de referință (numitorul) care îndeplinesc definiția bolii.

Rata *de incidență* este definită astfel:

$$\frac{\text{Numărul de cazuri noi ale unei boli pe unitatea de timp}}{\text{Numărul total de risc la începutul acestei perioade de timp}}$$

De exemplu, studiile au descoperit că incidența de zece ani a unui eveniment coronarian major (cum ar fi atacul de cord) în rândul bărbaților albi, cu vârsta cuprinsă între 30 și 59 de ani, cu tensiunea arterială diastolică de 105 mm Hg sau mai mult, la momentul în care au fost observați pentru prima dată, a fost de 183 la 1.000. [17] Aceasta înseamnă că dintre 1.000 de bărbați albi, cu vârsta cuprinsă între 30 și 59 de ani, care aveau tensiune arterială diastolică peste 105 mm Hg la începutul perioadei de observație de zece ani, 183 dintre ei au avut un eveniment coronarian major (infarct sau moarte subită) în următorii zece ani. În rândul bărbaților albi cu tensiune arterială diastolică <75 mm Hg, incidența pe zece ani a unui eveniment coronarian a fost de 55/1.000. Astfel, compararea acestor două rate de incidență, 183/1.000 pentru cei cu tensiune arterială mare față de 55/1.000 pentru cei cu tensiune arterială scăzută, poate duce la deducerea că tensiunea arterială crescută este un factor de risc pentru boala coronariană.

Adesea se poate auzi cuvântul „incidență" folosit atunci când ceea ce înseamnă cu adevărat este prevalență. Ar trebui să aveți grijă de o astfel de utilizare incorectă. De exemplu, s-ar putea să auzi sau chiar să citești într-un jurnal medical că incidența diabetului zaharat în 1973 a fost de 42,6 la 1.000 de persoane, cu vârste cuprinse între 45 și 64 de ani, când ceea ce se înțelege cu adevărat este că prevalența a fost de 42,6/1.000. Lucrul de reținut este că prevalența se referă la *existența unei boli* într-o anumită perioadă de timp, în timp ce incidența se referă la *cazuri noi* ale unei boli care se dezvoltă într-o anumită perioadă de timp.

Rețineți că *rata mortalității este o rată de incidență,* în timp ce *morbiditatea poate fi exprimată ca o rată de incidență sau prevalență.* Într-

o boală cronică, rata de prevalență este mai mare decât rata de incidență, deoarece prevalența include atât cazuri noi, cât și cazuri existente care ar fi putut să fi apărut pentru prima dată cu mult timp în urmă, dar pacienții afectați au continuat să trăiască cu această afecțiune. Pentru o boală care este fie rapid fatală, fie rapid vindecată, incidența și prevalența pot fi similare. Prevalența poate fi stabilită prin efectuarea unui sondaj sau a unui screening al unei populații țintă și numărarea cazurilor de boală existente la momentul anchetei. Acesta este un studiu transversal. Cifrele de incidență sunt mai greu de obținut decât cifrele de prevalență, deoarece pentru a stabili incidența este necesar să se identifice un grup de persoane libere de boala în cauză (cunoscută sub numele de cohortă), să le observe pe o perioadă de timp și să se determine câți dezvoltă boala în acea perioadă de timp. Implementarea unui astfel de proces este dificilă și costisitoare.

4.5 Rata de mortalitate standardizată

Rata *standardizată de mortalitate* (SMR) este raportul dintre numărul de decese observate și numărul de decese așteptate. Numărul așteptat pentru o anumită grupă de vârstă, de exemplu, este adesea obținut din statisticile populației.

$$SMR = \frac{\textit{decese observate}}{\textit{decese așteptate}}$$

4.6 Persoană-Ani de observație

Ocazional, se vede o rată prezentată ca un anumit număr de evenimente *pe persoană-ani de observație* , mai degrabă decât pe numărul de indivizi observați într-o anumită perioadă de timp. Pe persoană-ani (sau luni) este utilă ca unitate de măsură atunci când oamenii sunt observați pentru diferite perioade de timp. Să presupunem că observați cohorte de oameni fără boli de inimă pentru a determina dacă incidența bolilor de inimă este mai mare pentru fumători decât pentru cei care renunță. Persoanele care renunță trebuie definiți, de exemplu, ca fiind cei care renunță la mai mult de cinci ani înainte de începerea observației. S-ar putea defini diferit persoanele care

renunță și ar putea obține rezultate diferite, așa că este important să specificați definiția. Alte considerații includ controlul duratei de fumat, care ar fi o funcție de vârstă la începutul fumatului și vârsta la începutul perioadei de observare, numărul de țigări fumate și așa mai departe. Dar, pentru simplitate, vom presupune că toți dintre fumători au fumat o cantitate egală și toți cei care au renunțat au fumat o cantitate egală înainte de a renunța.

Putem exprima rata de incidență a bolilor de inimă într-o anumită unitate de timp, să zicem 10 ani, ca numărul care au dezvoltat boala în acel timp, împărțit la numărul de persoane pe care le-am observat (număr cu risc). Totuși, să presupunem că nu i-am observat pe toți pentru aceeași perioadă de timp. Acest lucru s-ar putea întâmpla deoarece unele persoane s-au mutat sau au murit din alte cauze sau au fost înscrise în studiu în momente diferite sau din alte motive. Într-un astfel de caz am putea folosi ca numitor numărul de *ani-persoană de observație.*

De exemplu, dacă individul 1 a fost înscris la momentul 0 și a fost observat timp de 4 ani, apoi a pierdut din urmărire, el ar fi contribuit cu 4 ani-persoană de observație. Zece astfel de indivizi ar contribui cu 40 de ani-persoană de observație. Un alt individ observat timp de 8 ani ar fi contribuit cu 8 persoană-ani de observație și 10 astfel de indivizi ar fi contribuit cu 80 persoană-ani de observație pentru un total de 120 persoană-ani. Dacă s-au dezvoltat șase cazuri de boli de inimă printre cei observați, rata ar fi de 6 la 120 de persoane-ani, mai degrabă decât de 6/10 indivizi observați. Rețineți că, dacă numitorul este dat în funcție de ani, nu știți dacă se referă la 120 de persoane observate fiecare timp de un an, sau la 12 persoane observate timp de 10 ani sau la o combinație. O altă problemă cu această metodă de exprimare a ratelor este că reflectă experiența medie de-a lungul intervalului de timp, dar este posibil ca rata bolilor de inimă să fie aceeași pentru fumători ca și pentru cei care renunță la fumat în primii 3 ani, iar ratele încep să se separe după aceea. În orice caz, sunt disponibile diverse metode statistice pentru a fi utilizate cu analiza persoană-an. O explicație excelentă a acestui subiect este dată în cartea *An Introduction to Epidemiologic Methods,* de Harold A. Kahn.

4.7 Variabile dependente și independente

În studiile de cercetare dorim să cuantificăm relația dintre un set de variabile, pe care le putem considera predictori sau determinanți, și o

variabilă de rezultat sau criteriu de care suntem interesați. Această variabilă de rezultat, pe care obiectivul nostru este să o explicăm, este variabila dependentă.

O *variabilă dependentă* este un factor a cărui valoare depinde de nivelul altui factor, care este denumit *variabilă independentă* . În exemplul - fumatului și mortalității prin cancer pulmonar, durata și/sau numărul de țigări fumate sunt variabile independente de care depinde mortalitatea prin cancer pulmonar (astfel, mortalitatea prin cancer pulmonar este variabila dependentă).

4.8 Tipuri de studii

În Secțiunea 1.4 am descris diferite tipuri de proiecte de studiu, în contextul discuției noastre despre metoda științifică și despre modul în care știm ceea ce știm. Acestea au fost studii observaționale, care pot fi transversale , caz-control sau prospective, și studii experimentale, care sunt studii clinice. În secțiunile următoare vom lua în considerare tipurile de inferențe care pot fi derivate din datele obținute din aceste modele diferite.

Obiectivul este de a evalua relația dintre un factor de interes (variabila independentă), pe care uneori îl vom numi expunere și o variabilă de rezultat (variabila dependentă).

Studiile observaționale se disting prin *momentul în care se fac măsurători pe variabilele dependente și independente* , așa cum este ilustrat mai jos. În studiile transversale, atât variabilele dependente, cât și cele independente (rezultat și expunere) sunt măsurate în același timp, în prezent. În studiile caz-control, rezultatul este măsurat acum și expunerea este estimată din trecut. În studiile prospective, expunerea (variabila independentă) este măsurată acum, iar rezultatul este măsurat în viitor. În secțiunea următoare vom discuta diferitele inferențe care trebuie făcute din studiile transversale versus studiile prospective.

	Timpul de măsurare		
	Trecut	**Prezent**	**Viitor**
Secțiune transversală:		rezultatul expunerii	

Controlul cazului:	expunere	rezultat	
Perspectivă:		expunere	rezultat

4.9 Analize transversale versus longitudinale asupra datelor

Studiile prospective sunt uneori cunoscute și ca studii longitudinale, deoarece oamenii sunt urmăriți longitudinal, în timp. Examinarea datelor longitudinale poate duce la inferențe destul de diferite decât cele care se obțin din analizele transversale ale datelor. De exemplu, luați în considerare vârsta și tensiunea arterială.

Studiile transversale au arătat în mod repetat că tensiunea arterială sistolica medie este mai mare în fiecare grup de vârstă succesiv de zece ani, în timp ce presiunea diastolică crește pentru grupele de vârstă până la 50 de ani și apoi atinge un platou. Din aceste tipuri de studii nu se poate spune că tensiunea arterială crește odată cu vârsta, deoarece presiunile măsurate pentru bărbații de 30 de ani, de exemplu, nu au fost obținute pe aceiași indivizi zece ani mai târziu, când aveau 40 de ani, ci au fost obținute pentru un set diferit de bărbați de 40 de ani. Pentru a determina efectul vârstei asupra tensiunii arteriale, ar trebui să aruncăm o privire longitudinală sau prospectivă asupra acelorași indivizi pe măsură ce îmbătrânesc. O interpretare a curbei observate pentru tensiunea arterială diastolică, de exemplu, ar putea fi aceea că indivizii de peste 60 de ani care aveau presiuni diastolice foarte mari au murit, lăsând în viață doar acei indivizi cu presiune mai scăzută suficient de mult pentru a fi incluși în eșantionul celor cărora li s-a măsurat tensiunea arterială în aspectul transversal.

Diagramele din Figura 4.2 ilustrează impactul posibil al unui „ efect cohort”, o vedere în secțiune transversală și o vedere longitudinală a acelorași date. (Literele indică grupuri de persoane examinate într-un anumit an.)

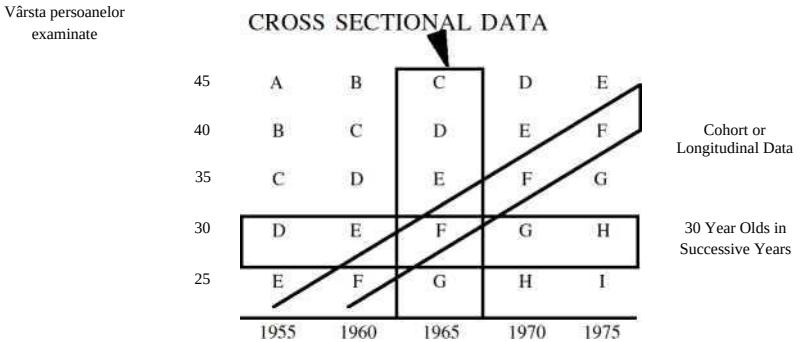


Figure 4.2

Dacă luați tensiunea arterială a tuturor grupelor în 1965 și comparați grupul F cu grupul D, veți avea o comparație transversală a persoanelor de 30 de ani cu persoanele de 40 de ani la un moment dat. Dacă comparați grupul F în 1965 cu grupul F (aceiași indivizi) în 1975, veți avea o comparație longitudinală. Dacă comparați grupul F în 1965 cu grupul H în 1975, veți avea o comparație a tensiunii arteriale la persoanele de 30 de ani în diferite momente în timp (o privire orizontală).

Aceste comparații pot duce la concluzii destul de diferite, așa cum se arată în exemplele schematice din Figura 4.3 folosind numere fictive pentru a reprezenta tensiunea arterială diastolică medie.

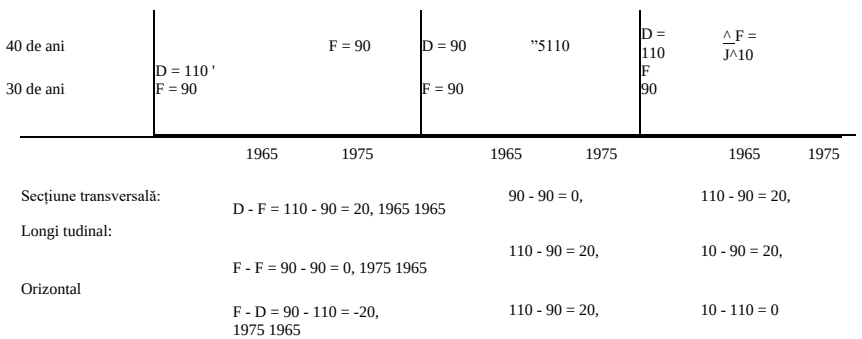


Figura 4.3

În exemplul (1), măsurătorile din 1965 indică faptul că tensiunea arterială diastolică medie pentru persoanele de 30 de ani (grupa F) a fost de 90 mm Hg, iar pentru persoanele de 40 de ani (grupa D), aceasta a fost de 110 mm Hg. Privind grupul F zece ani mai târziu, când aveau 40 de ani, indică tensiunea arterială diastolică medie a acestora a fost de 90 mm Hg. Rezultă următoarele calcule:

aspect transversal:	D - F = 110 - 90 = 20 1965 1965
concluzie:	Persoanele de 40 de ani au tensiune arterială mai mare decât cei de 30 de ani (cu 20 mm Hg).
aspect longitudinal:	F - F = 90 - 90 = 0 1975 1965
concluzie:	Tensiunea arterială nu crește odată cu vârsta.
aspect orizontal: (comparații de cohorte)	F - D = 90 - 110 = -20 1975 1965
concluzie:	Bătrânii de 40 de ani în 1975 au o tensiune arterială mai mică decât cei de 40 de ani în 1965

O posibilă interpretare : tensiunea arterială nu crește odată cu vârsta, dar pentru cohorta F au funcționat diferite forțe de mediu decât pentru cohorta D.

În exemplul (2) avem

aspect transversal:	D - F = 90 - 90 = 0 mm Hg 1965 1965
concluzie:	Din datele transversale, concluzionăm că tensiunea arterială nu este mai mare odată cu vârsta înaintată

aspect longitudinal: F - F = 110 - 90 = 20
1975 1965

concluzie: Din datele longitudinale concluzionăm că tensiunea arterială crește odată cu vârsta.

aspect orizontal: F - D = 110 - 90 = 20
1975 1965

concluzie: Oamenii de 40 de ani din 1975 au o tensiune arterială mai mare decât cei de 40 de ani din 1965

O posibilă interpretare : tensiunea arterială crește odată cu vârsta și diferiți factori de mediu operați pe cohorta F decât pe cohorta D.

În exemplul (3) avem

aspect transversal: D - F = 110 - 90 = 20
1965 1965

concluzie: La nivel transversal, a existat o creștere a tensiunii arteriale pentru persoanele de 40 de ani față de cea pentru cei de 30 de ani.

aspect longitudinal: F - F = 110 - 90 = 20
1975 1965

concluzie: Longitudinal se vede ca tensiunea arteriala creste odata cu inaintarea in varsta.

aspect orizontal: F - D = 110 - 110 = 0
1975 1965

concluzie: Nu a existat nicio modificare a tensiunii arteriale în rândul persoanelor de 40 de ani în perioada de zece ani.

O posibilă interpretare : tensiunea arterială crește odată cu vârsta (susținută

atât de date longitudinale, cât și de date transversale) și factorii de mediu afectează ambele cohorte în mod similar.

4.10 Măsuri ale riscului relativ: Inferențe din studii prospective: studiul Framingham

În studiile epidemiologice, suntem adesea interesați să știm cât de mult mai probabil este un individ să dezvolte o boală dacă este expus la un anumit factor decât individul care nu este atât de expus. O măsură simplă a unei astfel de probabilități se numește *risc relativ (RR).* Este raportul dintre două rate de incidență: *rata de dezvoltare a bolii pentru persoanele cu factor de expunere, împărțită la rata de dezvoltare a bolii pentru persoanele fără factor de expunere* . Să presupunem că dorim să determinăm efectul hipertensiunii arteriale (hipertensiunea) asupra dezvoltării bolilor cardiovasculare (BCV). Pentru a obține riscul relativ trebuie să calculăm ratele de incidență. Putem folosi datele dintr-un studiu prospectiv clasic, Framingham Heart Study. [18]

Acesta a fost un studiu epidemiologic prospectiv de pionierat al unui eșantion de populație din orășelul Framingham, Massachusetts. Începând cu 1948, o *cohortă* de oameni a fost selectată pentru a fi urmărită bienial . Termenul *de cohortă* se referă la un grup de indivizi urmăriți longitudinal pe o perioadă de timp. O cohortă de naștere, de exemplu, ar fi populația de indivizi născuți într-un anumit an. Cohorta Framingham a fost un eșantion de persoane alese la începutul perioadei de studiu și a inclus bărbați și femei cu vârsta cuprinsă între 30 și 62 de ani la începutul studiului. Acești indivizi au fost observați pe o perioadă de 20 de ani și au fost determinate morbiditatea și mortalitatea asociate bolilor cardiovasculare. S-a obținut o fișă standardizată a spitalului și un certificat de deces, examinarea clinică a fost repetată la intervale de doi ani, iar preocuparea majoră a studiului Framingham a fost evaluarea relației

dintre caracteristicile determinate la persoanele *în stare de sănătate* și dezvoltarea ulterioară a bolii.

Prin acest studiu au fost identificați „factori de risc" pentru bolile cardiovasculare. Factorii *de risc* sunt caracteristicile fiziologice antecedente sau obiceiurile alimentare și de viață, a căror prezență crește probabilitatea individului de a dezvolta boli cardiovasculare la un moment dat. Printre cei mai importanți factori predictivi identificați în studiul Framingham au fost *creșterea tensiunii arteriale, creșterea nivelului de colesterol seric și fumatul de țigară.* Glicemia crescută și rezultatele anormale ale electrocardiogramei de repaus sunt, de asemenea, predictive pentru viitoarele boli cardiovasculare.

Riscul relativ poate fi determinat prin următorul calcul:

$$\frac{\text{rata de incidență a bolilor cardiovasculare (cazuri noi) într-o anumită perioadă de timp în rândul persoanelor fără BCV la începutul perioadei de studiu care au factorul de risc în cauză (de exemplu, hipertensiune arterială)}}{\text{rata de incidență a BCV în perioada de timp dată în rândul persoanelor fără BCV inițial, care nu au factorul de risc în cauză (tensiune arterială normală)}}$$

$$\text{RR de CVD în interiorul 18 ani după prima examen} = \frac{\text{353,2/10.000 persoane cu cu hipertensiune arterială certă}}{\text{123,9/10.000 persoane cu risc fără hipertensiune}}$$

Din datele Framingham, $\frac{353.2}{123.9}$ calculăm pentru bărbați în studiu

Aceasta înseamnă că un bărbat cu hipertensiune arterială certă are de 2,85 ori mai multe șanse de a dezvolta BCV într-o perioadă de 18 ani decât un bărbat care nu are hipertensiune arterială. Pentru femei riscul relativ este

= 2 .
85

$$\frac{187.9}{57.3} = 3.28$$

Aceasta înseamnă că hipertensiunea arterială prezintă un risc relativ oarecum mai mare pentru femei. Dar rețineți că riscul *absolut* pentru persoanele cu hipertensiune arterială certă (adică, incidența BCV) este mai mare la bărbați decât la femei, fiind de 353,2 la 10.000 de bărbați față de 187,9 la 10.000 de femei.

Ratele de incidență prezentate mai sus au fost ajustate în funcție de vârstă. Ajustarea vârstei este discutată în Secțiunea 4.3. Adesea, cineva poate dori să se adapteze pentru alte variabile, cum ar fi starea de fumat, diabetul, nivelul de colesterol și alți factori care pot fi, de asemenea, legați de rezultat. Acest lucru poate fi realizat prin analiza de regresie logistică multiplă sau prin analiza pericolelor proporționale Cox, care sunt descrise în secțiunile 4.16 și 4.18, respectiv, dar mai întâi vom descrie modul în care riscul relativ poate fi calculat din studii prospective sau estimat din studii caz-control.

4.11 Calculul riscului relativ din studii prospective

Riscul relativ poate fi determinat direct din studii prospective prin construirea unui tabel 2 x 2 după cum urmează: [19]

		DISEASE (developing in the specified period): Yes	DISEASE (developing in the specified period): No	
RISK FACTOR (determined at beginning of study period)	PRESENT (high blood pressure)	$a = 90$	$b = 403$	$a + b = 493$ (persons with factor)
	ABSENT (normal blood pressure)	$c = 70$	$d = 1201$	$c + d = 1271$ (persons without factor)

Riscul relativ este

$$\frac{\text{incidența bolii în rândul celor cu TA mare}}{\text{incidența bolii în rândul celor cu TA normală}} = \frac{a/(a+b)}{c/(c+d)} \sim \frac{90/493}{70/1271} = 3.31$$

Riscul relativ sau raportul hazardului poate fi calculat din modelele de regresie a hazardelor proporționale Cox (care permit ajustarea pentru alte variabile), așa cum este descris în Secțiunea 4.19.

4.12 Odds Ratio: Estimarea riscului relativ din studiile de caz-control

Un studiu caz-control este cel în care investigatorul încearcă să stabilească o asociere între prezența unei caracteristici (un factor de risc) și apariția unei boli, *pornind de la un eșantion de persoane bolnave și un grup de control de persoane nebolnave și notând prezența sau absența caracteristicii în fiecare dintre aceste două grupuri.* Poate fi ilustrat prin construirea unui tabel 2 x 2 după cum urmează:

		BOALA: PREZENT	BOALA: ABSENT
RISC FACTOR	PREZENT	*o*	*b*
	ABSENT	*c*	*d*
		a + c (numar de persoane *cu* boala)	*b + d* (numar de persoane *fara* boala)

Obiectivul este de a determina dacă proporția persoanelor cu boală care au factorul este mai mare decât proporția persoanelor fără boală care au factorul. Cu alte cuvinte, se dorește să știm dacă $a/(a + c)$ este mai mare decât $b/(b + d)$.

Studiile caz-control sunt adesea denumite *studii retrospective* deoarece investigatorul trebuie să adune date despre variabilele *independente retrospectiv.* Variabila dependentă – prezența bolii – este obținută în

momentul prelevării, spre deosebire de studiile prospective în care variabilele independente sunt măsurate în momentul prelevării, iar variabila dependentă este măsurată într-un moment viitor (adică, când boala se dezvoltă). Adevărata distincție între studiile caz-control sau retrospective și studiile prospective are de-a face cu selectarea indivizilor pentru studiu - cei *cu și fără boală* în studiile caz-control/retrospective și cei *cu și fără factorul* de interes în studiile prospective.

Deoarece în studiile prospective *eșantionăm persoanele cu caracteristica de interes și persoanele fără caracteristica,* putem obține riscul relativ *direct* prin calcularea ratelor de incidență a bolii în aceste două grupuri. În studiile caz-control, totuși, *eșantionăm pacienți cu și fără boală* și notăm prezența sau absența caracteristicii de interes în aceste două grupuri; nu avem o măsură directă a *incidenței* bolii. Cu toate acestea, luând anumite ipoteze, putem face câteva aproximări ale riscului relativ dacă am putea măsura ratele de incidență printr-un studiu prospectiv. Aceste aproximări sunt valabile cel mai bine pentru bolile cu *incidență scăzută.*

Pentru a estima riscul relativ din studiile caz-control, rețineți că

$$a/(a + b) \_ a(c + d)$$
$$c/(c + d) c (a + b)$$

Acum presupuneți că, deoarece boala nu este atât de comună, *c* este - neglijabilă în raport cu *d* (cu alte cuvinte, printre persoanele fără factor de risc nu sunt atât de mulți care vor face boala, raportat la numărul de persoane care nu o vor face). Să presupunem, de asemenea, că, *în populație, a* este neglijabil în raport cu *b* , deoarece chiar și printre persoanele cu factor de risc nu atât de mulți vor face boala în comparație cu numărul celor care nu o vor face. Atunci termenii din paranteze devin *d* la numărător și *b* la numitor, astfel încât

$$\frac{a\,(c+d)}{c\,(a+b)} \text{ se reduce la } OR = \frac{ad}{bc}$$

Aceasta este cunoscută sub numele de *odds ratio* (OR) și este o bună estimare a riscului relativ atunci când boala este rară.

Un exemplu despre modul în care este calculat cota de cote este prezentat mai jos. Într-un studiu caz-control al cancerului pulmonar a fost obținut tabelul de mai jos. [20] Rețineți că aici nu eșantionăm fumători și nefumători. Mai degrabă le eșantionăm pe cei cu și fără boală. Deci, deși în *populația generală a* este mic în raport cu *b* , în acest eșantion nu este așa.

	Patients with Lung Cancer		Matched Controls with Other Diseases	
Smokers of 15-24 cigarettes daily	475	*a*	431	*b*
Nonsmokers	7	*c*	61	*d*
	(persons with disease)		(persons without disease)	

Odds ratio, ca estimare a riscului relativ de a dezvolta cancer pulmonar pentru persoanele care fumează 15-24 de țigări pe zi, în comparație cu nefumătorii este

$$\textit{Odds ratio} = \frac{475 \times 61}{431 \times 7} = 9.60 = \textit{estimate of relative risk}$$

Aceasta înseamnă că fumătorii de 15-24 de țigări pe zi au de 9,6 ori mai multe șanse de a face cancer pulmonar decât nefumătorii.

Încă un lucru despre odds ratio: este raportul dintre șansele de cancer pulmonar pentru cei care fumează 15-24 de țigări pe zi, comparativ cu șansele de cancer pulmonar pentru cei care nu fumează. În exemplul de mai sus, obținem

pentru fumători: șansele de cancer pulmonar sunt pentru nefumători: șansele de cancer pulmonar sunt $\frac{475}{431}$

Deci, ideea este că raportul de șanse este raportul de $\frac{7}{61}$

$$ratio\ of\ odds = \frac{475 / 431}{7 / 61}$$

șanse, indiferent dacă boala este rară sau nu. Este întotdeauna raportul dintre șansele de îmbolnăvire pentru cei cu expunere față de șansele de îmbolnăvire pentru cei fără expunere. Dar atunci când boala este rară, este și o bună estimare a riscului relativ. Putem, de asemenea, să punem limite de încredere pentru raportul de cote, prezentat în Anexa F. Ratele de cote pot fi calculate din regresia logistică (care permite ajustarea pentru alte variabile), așa cum este descris în Secțiunea 4.17.

4.13 Risc atribuibil

Riscul atribuibil (RA) este:

risk in exposed — risc la indivizi neexpusi.

Riscul atribuibil populației (PAR) este:

RA x prevalența factorului de risc

În timp ce riscul relativ se referă la riscul unei boli la persoanele expuse *în raport cu riscul* la persoanele neexpuse, riscul atribuibil se referă la *diferența de risc absolut* al persoanelor expuse comparativ cu persoanele neexpuse. Ne poate spune cât de mult *exces de risc* există din cauza expunerii la cei expuși. În exemplul din Secțiunea 4.11, riscul de 10 ani în rândul celor cu tensiune arterială mare a fost 90/493 = 0,183, (sau 183 la 1000 de persoane cu tensiune arterială mare), în timp ce la cei cu tensiune arterială normală a fost 70/1271 = 0,055 (sau 55 la 1000 cu presiune normală).

Astfel, riscul atribuibil la cei expuși (adică cu hipertensiune arterială) este 0,183-0,055 = 0,128 (128 la 1000). Cu alte cuvinte, evenimentele de boli de inimă la 128 din 183 de persoane la 1000 cu hipertensiune arterială pot fi atribuite hipertensiunii arteriale. De asemenea, putem exprima acest exces ca procent din riscul expusului care este atribuit expunerii:

$$\frac{128\ /1000}{183\ /1000} = 128 = .70\ \textit{or}\ 70\%$$

DAR, trebuie să fim foarte atenți cu privire la o astfel de atribuire - este valabilă numai atunci când putem presupune că expunerea este cauza bolii (după ce luăm în considerare factorii de confuzie și alte surse de părtinire).

Riscul atribuibil populației (PAR) este o măsură utilă atunci când dorim să vedem cum am putea reduce morbiditatea sau mortalitatea prin eliminarea unui factor de risc. Depinde de prevalența factorului de risc în populație, așa cum sa menționat mai sus. Iată un exemplu din Inițiativa pentru sănătatea femeilor (descris mai detaliat în capitolul 6). S-a descoperit într-un studiu clinic că femeile aflate în postmenopauză care luau estrogen plus progestativ au avut o rată anuală a bolii coronariene de 39 la 10.000, comparativ cu o rată de 33 la 10.000 pentru femeile care luau placebo. [21]

Astfel:

$$AR = \frac{39 - 33}{10.000} = .0006$$

sau 6 evenimente de boală coronariană în exces la 10.000 de femei care iau acest preparat, pe an.

Deoarece aproximativ 6.000.000 de femei luau acel preparat hormonal la momentul respectiv (expuse), atunci .0007 x 6.000.000 = 3600 de evenimente coronare de boli cardiace pe an ar putea fi *atribuite* luării de estrogen plus progestativ.

Prevalența utilizării estrogenului plus progestativ estimată din același studiu când a fost început pentru prima dată a fost de aproximativ 18%. Dacă folosim această estimare,

PAR = AR x prevalența factorului de risc = .0006 x .18 = .000108;

Astfel, dacă s-ar elimina utilizarea de estrogen plus progestativ, ar exista 10,8 la 100.000 de femei în postmenopauză care au avut mai puține evenimente de boală coronariană.

4.14 Bias de răspuns

Există multe tipuri diferite de părtinire care v-ar putea conduce fie să - subestimați, fie să supraestimați dimensiunea unui raport de risc relativ al cotelor și este important să încercați să anticipați sursele potențiale de părtinire și să le evitați. Ilustrația de pe pagina următoare arată impactul unui tip de posibilă prejudecată: *prejudecata de constatare sau de răspuns.*

Să presupunem că aveți următoarea situație. Din 100 de persoane expuse unui factor de risc, 20% dezvoltă boala și din 100 de persoane neexpuse , 16% dezvoltă boala, rezultând un risc relativ de 1,25, așa cum se arată în ilustrație.

Acum imaginați-vă că doar 60% dintre cei expuși răspund la urmărire sau sunt constatați că au sau nu au boala, *o rată de răspuns de 60% în rândul celor expuși* . Mai presupunem că toți cei care nu răspund se întâmplă să fie printre cei care *nu* dezvoltă boală. Riscul relativ ar fi calculat ca 2,06.

Acum imaginați-vă că doar 60% din răspunsul neexpus, o *rată de răspuns de 60% în rândul celor neexpuși* și toți cei neexpuși care nu răspund se întâmplă să fie printre cei care nu au boala . Acum riscul relativ estimat este de 0,75.

Pentru a rezuma, puteți obține estimări contradictorii ale riscului relativ dacă aveți rate de răspuns diferențiate. Prin urmare, este foarte important să obțineți un răspuns sau o constatare cât mai completă posibil. Urmează tabelele care prezintă aceste calcule.

RATA COMPLETĂ DE RĂSPUNS

EXPOSURE	BOALA +	BOALA -	
+	R = 100% 20	100% 80	100
-	100% 16	100% 84	100
	36	164	200

$$RR = \frac{20 / 100}{16 / 100} = \frac{20}{16} = 1$$

DIFERENȚIAL RATA DE RĂSPUNS

EXPOSURE	BOALA +	BOALA -	
+	R = 100% 20	50% 40	60 (rata de răspuns = 60%)
	100% 16	100% 84	100
	36	124	160

$$RR = \frac{20 / 60}{16 / 100} = \frac{33}{16} = 1$$

DIFERENŢIAL RATA DE RĂSPUNS

EXPOSURE	BOALA +	BOALA -	
+	R = 100% 20	100% 40	100
	100% 16	52% 84	60 (rata de răspuns = 60%)
	36	124	160

$$RR = \frac{20 / 100}{16 / 60} = \frac{\bullet 20}{\bullet 266} = \text{---}$$

4.15 Variabile de confuzie

O *variabilă de confuzie* este una care este strâns asociată atât cu variabila independentă, cât și cu rezultatul de interes pentru cei neexpuși. De exemplu, o variabilă de confuzie în studiile privind cafeaua și bolile de inimă poate fi fumatul. Deoarece unii băutori de cafea sunt și fumători, dacă un studiu a găsit o relație între consumul de cafea (variabila independentă) și dezvoltarea bolilor de inimă (variabila dependentă), ar putea însemna cu adevărat că fumatul este cel care provoacă boli de inimă, mai degrabă decât cafeaua. În acest exemplu, fumatul este variabila de confuzie.

Dacă *atât* variabila de confuzie, cât și variabila independentă de interes sunt strâns asociate cu variabila dependentă, atunci relația observată dintre variabilele independente și dependente poate fi într-adevăr o reflectare a relației *adevărate* dintre variabila confundatoare și variabila dependentă. O modalitate intuitivă de a privi acest lucru este să ne imaginăm că, dacă o variabilă de confuzie ar fi perfect asociată cu o variabilă independentă, aceasta ar putea fi înlocuită. Este important să se țină seama sau să se ajusteze pentru variabilele confuze *în proiectarea și analiza statistică a*

studiilor pentru a evita inferențe greșite.

Există mai multe abordări pentru a trata potențialii factori de confuzie. O abordare este de a trata aceasta în proiectarea studiului prin potrivire, de exemplu, așa cum este descris în Secțiunea 4.16 de mai jos; un alt mod de a controla variabilele confuze este în faza de analiză a datelor, prin utilizarea analizei cu mai multe variante, așa cum este descris în Secțiunile 4.17, 4.18 și 4.20 de mai jos. O discuție excelentă se găsește în *Modern Epidemiology* de Kenneth J. Rothman și Sander Greenland.

4.16 Potrivire

O abordare pentru a trata potențialii factori de confuzie este de a potrivi subiecții din cele două grupuri pe variabila de confuzie. În exemplul discutat mai sus cu privire la studiile despre cafea și bolile de inimă, am putea potrivi subiecții în istoricul lor de fumat, deoarece fumatul poate fi un factor de confuzie al relației dintre cafea și bolile de inimă. Ori de câte ori am înscris un băutor de cafea în studiu, ne-am determina dacă acea persoană era fumătoare. Dacă pacientul a fost un fumător, următorul pacient care ar fi înrolat și care nu a fost un băutor de cafea (adică, un membru al grupului de comparație), ar trebui să fie și el fumător. Pentru fiecare nefumător care consumă cafea, va fi înscris un nefumător care nu consumă cafea . În acest fel, am avea același număr de fumători în cele două grupuri. Aceasta este cunoscută sub numele de *potrivire unu-la-unu.* Există și alte modalități de potrivire și acestea sunt discutate mai pe larg în secțiunea Lecturi sugerate, în special în *Metode statistice pentru studii comparative* de Anderson și colab. și în *Relații cauzale în medicină* de J. Mark Elwood.

În studiile caz-control, găsirea unui grup de comparație adecvat poate fi dificilă. De exemplu, să presupunem că un investigator studiază efectul cafelei asupra cancerului pancreatic. Anchetatorul alege ca controale , pacienți aflați în spital în același timp și în aceeași secție cu cazurile, dar cu alt diagnostic decât cancerul. Este posibil ca pacienții internați pentru alte probleme gastrointestinale decât cancerul să fi renunțat voluntar la consumul de cafea, deoarece le-a deranjat stomacul. Într-o astfel de situație, obiceiurile de consum de cafea ale celor două grupuri ar putea fi similare și investigatorul ar putea să nu găsească o asociere mai mare a consumului de

cafea cu cazurile decât cu martorii. Un grup mai adecvat ar putea fi pacienții dintr-o secție diferită, să zicem o secție de ortopedie. Dar aici ar trebui să fie atenți să se potrivească în funcție de vârstă, deoarece pacienții ortopedici pot fi mai tineri decât în celelalte cazuri dacă spitalul se întâmplă să fie într-o zonă de schi, de exemplu, unde schiul nesăbuit duce la picioare rupte, sau pot fi substanțial mai în vârstă decât în celelalte cazuri dacă există mulți pacienți cu proteze de șold din cauza căderilor la vârstnici sau a osteoartritei.

Trebuie subliniat că factorul care este potrivit nu poate fi evaluat în ceea ce privește relația cu rezultatul. Astfel, dacă comparăm două grupuri de femei pentru efectul aportului de vitamina A asupra cancerului de col uterin și facem un studiu caz-control în care înrolăm cazuri de cancer de col uterin și controale potrivite pe vârstă, nu vom putea spune din acest studiu dacă vârsta este legată de cancerul de col uterin. Acest lucru se datorează faptului că ne-am asigurat că distribuțiile de vârstă sunt aceleași atât în grupul de caz, cât și în grupul de control, prin potrivirea pe vârstă, deci evident că nu vom putea găsi diferențe de vârstă între grupuri.

Unii statisticieni consideră că potrivirea se face adesea inutil și că, dacă aveți un studiu suficient de mare, randomizarea simplă sau randomizarea stratificată este adecvată pentru a asigura o distribuție echilibrată a factorilor de confuzie. În plus, metodele de analiză multivariată, cum ar fi regresia logistică sau modelele de pericole proporționale, oferă o altă modalitate, de obicei mai bună, de a controla factorii de confuzie. O bună discuție despre potrivire poate fi găsită în cartea *Methods in Observational Epidemiology* , de Kelsey, Thompson și Evans.

4.17 Regresie logistică multiplă

Analiza de regresie logistică multiplă este utilizată pentru a calcula probabilitatea ca un eveniment să se întâmple în funcție de mai multe variabile independente. Este util în controlul factorilor de confuzie atunci când se examinează relația dintre o variabilă independentă și apariția unei

rezultat (de exemplu, cum ar fi atacul de cord) într-o anumită perioadă de timp.
Ecuația ia forma

$$P(\text{eveniment}) = \frac{1}{1 + e^{\sim k}}$$

$$\text{unde } k \text{ — } C_O + C_1 X_1 + C_2 X_2 + C_3 X_3 + \ldots + C_m X_m$$

Fiecare X_i este o anumită variabilă independentă și coeficienții corespunzători, C 's, sunt calculați din datele obținute în studiu. De exemplu, să luăm datele Framingham pentru probabilitatea de a dezvolta o boală cardiovasculară în decurs de 8 ani. Boala cardiovasculară (BCV) a fost definită ca boală coronariană, infarct cerebral, claudicație intermitentă sau insuficiență cardiacă congestivă.

$$P(CVD) = \frac{1}{1 + e^{-[-19.77 + 37(\text{vârsta}) - .002(\text{vârsta})^2 + .026(\text{chl}) + .016(\text{SBP}) + .558(\text{SM}) + 1.053(\text{LVH}) + .602(\text{Gl}) - .002(\text{Gl}) - .00036)](\text{chl x vârsta})}}$$

unde chl = colesterol seric,
SBP = tensiunea arterială sistolica,
SM = 1 dacă da pentru fumat, 0 dacă nu,
LVH = hipertrofie ventriculară stângă, 1 dacă da, 0 dacă nu,
Gl = intoleranță la glucoză, 1 dacă da, 0 dacă nu.

De exemplu, să presupunem că luăm în considerare un bărbat de 50 de ani al cărui colesterol este 200, tensiunea arterială sistolică este 160, care fumează, nu are LVH și nu are intoleranță la glucoză. Când înmulțim coeficienții cu valorile acestui individ asupra variabilelor independente și facem calculele necesare, obținem o probabilitate de .17. Aceasta înseamnă că acest individ are 17 șanse din 100 de a dezvolta o formă de boală cardio-vasculară în următorii 8 ani.

Coeficienții dintr-o analiză de regresie logistică multiplă pot fi utilizați pentru a calcula raportul de șanse pentru un factor, controlând în același timp pentru toți

ceilalți factori. Modalitatea de a face acest lucru este să luăm logaritmul natural e ridicat la coeficientul pentru variabila de interes, dacă variabila este una dihotomică (adică, codificată ca 1 sau 0). De exemplu, șansele de apariție a bolilor cardiovasculare pentru fumători în comparație cu nefumătorii în rândul bărbaților, controlând în același timp vârsta, colesterolul, tensiunea arterială sistolica, hipertrofia ventriculară stângă și intoleranța la glucoză este de $_{558}{}^{=}$ 1,75. Aceasta înseamnă că o persoană care fumează are un risc de 1,75 ori mai mare de a face BCV (în decurs de 8 ani) decât cel care nu fumează dacă acești doi indivizi sunt egali în raport cu celelalte variabile din ecuație. Acest lucru este echivalent cu a spune că riscul fumătorului este cu 75% mai mare decât al nefumătorului .

Dacă vrem să comparăm șansele unei persoane cu o tensiune arterială sistolica de 200 față de cineva cu tensiunea arterială sistolica de 120, toți ceilalți factori fiind egali, o calculăm după cum urmează:

$$SAU = {}_{g}V_{(200 - 120)}{}_{-}{}_{g}{}_{.016(80)}{}_{-}{}_{g}{}_{1,28}{}_{-}$$

Bărbatul cu tensiune arterială sistolică de 200 mm Hg are de 3,6 ori mai multe șanse de a dezvolta boală decât cel cu presiune de 120. (Bineînțeles, acest lucru ar implica că cineva cu tensiune arterială sistolică de 260 ar fi, de asemenea, de 3,6 ori mai probabil să dezvolte BCV decât unul cu presiunea de 180. Dacă presupunerea unei creșteri liniare a riscului ar fi incorectă).

Regresia logistică poate fi folosită și pentru studii caz-control. În acest caz, ridicarea *lui g* la coeficientul variabilei de interes ne oferă și raportul cotelor, dar nu putem folosi ecuația pentru a prezice probabilitatea unui eveniment, deoarece am eșantionat din cazuri și controale, nu dintr-o populație generală de interes.

Analiza de regresie logistică multiplă a devenit utilizată pe scară largă, în mare parte datorită apariției computerelor de mare viteză, deoarece calcularea coeficienților necesită o mare putere de calculator. Pachetele statistice sunt disponibile pentru computerele personale.

Regresia logistică multiplă este adecvată atunci când variabila dependentă (rezultatul) este dihotomică (adică poate fi codificată ca 1 = eveniment, 0 = niciun eveniment) și când întrebarea se referă la apariția evenimentului de interes într-o anumită perioadă de timp și toți oamenii

sunt urmăriți pentru acea perioadă de timp. Cu toate acestea, atunci când timpul de urmărire pentru persoanele din studiu diferă, trebuie utilizată analiza de supraviețuire, așa cum este descris în secțiunile 4.19 și 4.20.

4.18 Confuzie prin indicație

Un tip de confuzie care poate apărea în studiile observaționale atunci când vă uitați la efectele tratamentului medicamentos asupra evenimentelor viitoare, este *confuzia prin indicație.* De exemplu, să presupunem că doriți să comparați efectele asupra bolilor de inimă ale diferitelor medicamente pentru hipertensiune arterială. Stabiliți ce medicamente antihipertensive iau participanții la studiu la o examinare inițială și apoi îi urmăriți în timp pentru a vedea cine dezvoltă boli de inimă. Problema poate fi că motivul pentru care oamenii luau diferite medicamente la momentul inițial a fost că aveau indicații diferite pentru ele și medicii le-au prescris medicamente adecvate acelor indicații. Astfel, persoanelor cu boală renală li s-ar putea să li se fi prescris un alt medicament pentru a controla tensiunea arterială crescută decât celor cu angină pectorală sau celor fără alte afecțiuni medicale, iar fiecare dintre aceste indicații poate fi legată diferit de rezultatul bolilor de inimă. Doar un studiu clinic, în care pacienții sunt repartizați aleatoriu pentru fiecare tratament medicamentos, poate răspunde cu adevărat la întrebarea despre diferitele efecte ale medicamentelor.

Cu toate acestea, există modalități de a minimiza confuzia prin indicație în studiile observaționale; o modalitate este de a exclude din analiză persoanele care au angină sau boli de rinichi din exemplul de mai sus. O altă metodă care se folosește este *analiza propensității.* [11] Ideea generală este că preziceți cine este probabil să ia medicamentul din variabilele independente pe care le-ați măsurat și să calculați un indice de „propensitate" pentru a lua medicamentul. Scorul de înclinație este apoi introdus ca o variabilă independentă în ecuația multivariată finală, împreună cu un subset de variabile pentru care controlați. Datele fiecărei persoane includ valorile covariatelor și scorul său de propensiune.

De exemplu, în exemplul hipertensiunii arteriale, dacă căutăm să vedem dacă un blocant al canalelor de calciu este asociat cu mortalitatea, dorim să ținem cont de faptul că, într-un studiu observațional, oamenii ar

putea fi mai probabil să li se prescrie un blocant al canalelor de calciu dacă au avut angină, de exemplu, și știm că angina este legată de mortalitate . Apoi am putea face următorii pași:

(1) Calculați o regresie logistică multiplă în care:
Y=1 dacă se administrează medicament, 0 în caz contrar (variabilă dependentă)
X (variabile independente) = vârsta, rasa/etnia, angina pectorală, IMC, tensiunea arterială sistolică, alte covariabile care ar putea influența prescrierea unui blocant al canalelor de calciu.
(2) Calculați un scor de înclinație pentru fiecare persoană (un indice bazat pe regresia dezvoltată mai sus)
(3) Calculați regresia care vă interesează cu adevărat care este să determinați asocierea blocanților canalelor de calciu cu mortalitatea după controlul potențialilor factori de confuzie, unde:
Z=1, dacă eveniment mortal, 0 în caz contrar (variabilă dependentă)
X = scorul de înclinație, vârsta, rasa/etnia, plus unele dintre celelalte covariabile relevante care se aflau în ecuația originală a înclinației.

O problemă cu analiza înclinației este că depinde de ce variabile trebuie să introduceți în ecuație pentru a obține scorul de propensiune; pot descrie înclinația sau nu. Există opinii diferite cu privire la cât de utilă este o astfel de analiză. Este descris mai detaliat în referințele date la sfârșit.

4.19 Analiza supraviețuirii: Metode de tabel de viață

Analiza de supraviețuire a datelor ar trebui să fie utilizată atunci când perioadele de urmărire diferă mult pentru diferite persoane sau când aceștia intră în studiu în momente diferite. Poate deveni destul de complex și această secțiune este destinată doar introducerii conceptelor. Să presupunem că doriți să comparați supraviețuirea pacienților tratați prin două metode diferite și să presupunem că aveți datele prezentate mai jos. [23] O vom analiza folosind curbele de supraviețuire Kaplan-Meier.

DECES LA O LUNĂ DATE ÎN DOUĂ GRUPE

Stare: (D = decedat în luna respectivă; L = trăind în luna respectivă).

(+ înseamnă că pacientul a fost pierdut de urmărire și a fost văzut ultima dată în viață în luna respectivă.)

Stare:	DLDDD	DDDDL
Grupa A:	4, 5+, 9, 11, 12	Grupa B: 2, 3, 4, 5, 6+

În fiecare grup, patru pacienți au murit până la 12 luni, iar unul a fost văzut în viață ceva timp în acel an, așa că nu știm dacă acel pacient era mort sau în viață la sfârșitul anului. Dacă am privi datele în acest fel, ar trebui să spunem că supraviețuirea cu un an a fost aceeași în ambele grupuri.

	Grup	
La sfarsitul a 12 luni	**O**	**B**
Mort	4	4
În	1	1
Rata de supraviețuire	20%	20%

Cu toate acestea, o modalitate mai adecvată de a analiza astfel de date este prin *curbele de supraviețuire.* Punctele pentru curbe sunt calculate așa cum se arată în tabelul de mai jos.

Col.1	Col.2	Col.3	Col.4	Col.5	Col.6	Col.7
Caz#	Ora în Mos.	Stare	# puncte. Intră	Prop. *qi mort* = = mort = introdus	Prop. Surv. P_1 = ^{1} - q_t	Cum. Surv. P_1 = P_{i-1} x P_t
Grupa A						
1	4	mort	5	1/5 = 0,2	0,80	1 x .8 = .8
2	5	surv	4	0/4 = 0,0	1.00	.8 x 1 = .8
3	9	mort	3	1/3 = 0,33	0,67	.8 x .67 = .53
4	11	mort	2	1/2 = 0,5	0,50	.53 x .5 = 27
5	12	mort	1	1/1 = 1,0	0,00	.27 x 0 = 0

Col.1	Col.2	Col.3	Col.4	Col.5	Col.6	Col.7
Caz#	Ora în Mos.	Stare	# puncte. Intră	Prop. Mort q = = mort = introdus	Prop. Surv. P_i = $_i$-q $_i$	Cum. Surv. P_1 = P_{-1} x P_i
Grupa B						
1	2	mort	5	1/5 = 0,2	0,80	$_i$ x .8 = .8
2	3	mort	4	1/4 = 0,25	0,75	.8 $_x$ .75 = .6
3	4	mort	3	1/3 = 0,33	0,67	.6 $_x$ .67 = .4
4	5	mort	2	1/2 = 0,5	0,50	.4 $_x$ .5 = .2
5	6	surv	1	0/1 = 0,0	1.00	.2 x . $_i$ = .2

În primul rând, pacienții sunt așezați în ordinea orei morții lor sau a ultimei viziuni în care au fost văzuți în viață. Să trecem prin al treilea rând pentru grupa A. ca exemplu.

Al treilea pacient a murit la 9 luni (coloanele 1 și 2). La începutul lunii a 9-a erau trei pacienți cu risc de deces (din totalul de cinci pacienți care au intrat în studiu). Acest lucru se datorează faptului că unul dintre cei cinci pacienți murise deja în luna a 4-a (cazul #1),

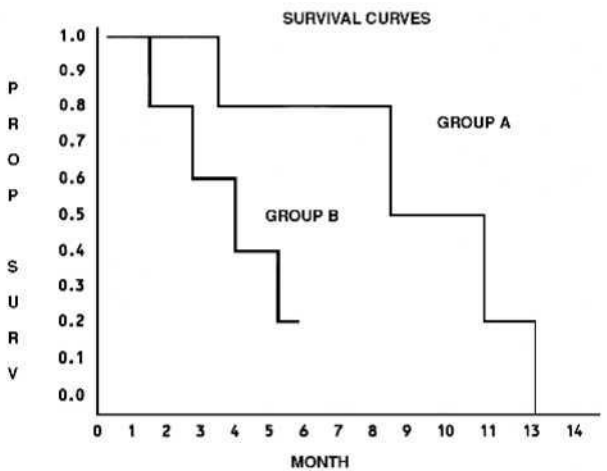


Figura 4.4

iar unul a fost văzut ultima dată în viață în luna a 5-a (cazul #2) și, prin

urmare, nu a fost disponibil pentru a fi observat. Dintre acești trei pacienți cu risc la începutul lunii a 9-a, unul a murit (cazul nr. 3). Deci probabilitatea de a muri în luna a 9-a este 1/3 și numim acest lucru q_t unde *i* în acest caz se referă la luna a 9-a. Prin urmare, proporția care supraviețuiește în luna a 9-a este $p_t = 1 - q. = 1 - .33 = .67$.

Proporția cumulativă de supraviețuire înseamnă proporția de - supraviețuire până în luna a 9-a. Pentru a supraviețui până în luna a 9-a, un pacient trebuia să fi supraviețuit până la sfârșitul lunii 8 *și* să fi supraviețuit în luna 9. Astfel, este egală cu probabilitatea cumulată de a supraviețui *până în* luna a 9-a, care este 0,8, din coloana 7 rândul 2, *și* de a supraviețui în luna a 9-a, care este 0,67. Înmulțim aceste probabilități pentru a obține .8 x .67 = .53 ca probabilitate de supraviețuire până în luna a 9-a. Dacă reprezentăm aceste puncte ca în Figura 4.4, observăm că cele două curbe de supraviețuire arată destul de diferit și că grupul A s-a descurcat mult mai bine.

Analiza supraviețuirii devine mai complicată atunci când presupunem că pacienții care au fost pierduți de urmărire într-un anumit interval de timp ar fi murit în același ritm ca acei pacienți despre care am avut informații . Alternativ, putem face calculele presupunând că toți au murit în intervalul în care au fost pierdute de urmărire sau că toți au supraviețuit în acel interval.

Analiza supraviețuirii poate fi făcută și în timp ce se controlează - variabilele de confuzie, folosind modelul hazardelor proporționale Cox.

4.20 Modelul de pericole proporționale Cox

Modelul de riscuri proporționale Cox este o formă de analiză multivariată a supraviețuirii care poate controla alți factori. *Variabila dependentă este timpul până la eveniment (sau timpul de supraviețuire),* care ar putea fi moartea, atacul de cord sau orice alt eveniment de interes. Acest lucru este în contrast cu regresia logistică multiplă , unde variabila dependentă este o variabilă da sau nu.

Modelul de riscuri proporționale Cox este utilizat în mod corespunzător atunci când există perioade de urmărire diferite, deoarece unii oameni s-au retras din studiu sau nu pot fi contactați. Persoanele care se încadrează într-una dintre aceste categorii sunt considerate a avea observații „cenzurate”.

Dacă evenimentul de interes este, să zicem, un accident vascular cerebral, atunci și persoanele care au murit în timpul studiului din cauze accidentale ar fi, de asemenea, „cenzurate” pentru că nu am putea ști dacă ar fi suferit sau nu un accident vascular cerebral, dacă ar fi trăit până la sfârșitul studiului.

Coeficienții din această analiză pot fi utilizați pentru a calcula riscul relativ de eveniment, după controlul celorlalte covariate din ecuație. Riscul relativ din modelele de pericole proporționale Cox este mai precis numit „raportul de risc”, dar va fi denumit aici risc relativ de dragul simplificării explicațiilor. Un exemplu de interpretare a rezultatelor unei astfel de analize este dat din Programul de hipertensiune sistolică la vârstnici (SHEP). Acesta a fost un studiu pe 4.736 de persoane de peste 60 de ani cu hipertensiune sistolica izolata (de exemplu, persoane cu tensiune arteriala sistolica mare si tensiune arteriala diastolica normala) pentru a vedea daca tratamentul cu doze mici de diuretic si/sau beta- blocant ar reduce rata accidentelor vasculare cerebrale comparativ cu rata din grupul de control tratat cu placebo.

Un eșantion de imprimare parțială computerizată a unei analize de regresie Cox din studiul SHEP este prezentat mai jos. Evenimentul de interes este accidentul vascular cerebral în grupul placebo.

Variabilă independentă	Coeficientul beta	SE	e^{Beta} = RR
Rasă	-0,1031	.26070	0,90
Sex (masculin)	0,1707	.19520	1.19
Vârstă	0,0598	.01405	1.06
Istoric de diabet	0,5322	.23970	1,70
Fumatul (nivel de referință)	0,6214	.23900	1,86

Să ne uităm la istoria diabetului. RR = e'^{5322} = 1,70, care este logaritmul natural *e* ridicat la puterea specificată de coeficientul beta ; *e* = 2,7183. (Nu întrebați de ce.) Aceasta înseamnă că o persoană cu hipertensiune sistolică netratată, care are antecedente de diabet, are un risc de 1,7 ori mai mare de a avea un accident vascular cerebral decât o persoană cu aceleași alte caracteristici, dar fără diabet. Acest lucru poate fi, de asemenea, declarat ca

un risc cu 70% mai mare. Limitele de încredere de 95% pentru riscul relativ sunt 1,06, 2,72, ceea ce înseamnă că suntem 95% încrezători că riscul relativ de accident vascular cerebral pentru cei cu antecedente de diabet se află în intervalul cuprins între 1,06, 2,72. Formula pentru intervalul de încredere de 95% pentru riscul relativ este

$$\text{Limita 1} = e^{[\text{beta} - 1.96(S.R)]}$$

$$\text{Limita 2} = e^{[\text{beta} + 1.96(S.R)]}$$

Dacă avem de-a face cu o variabilă continuă, cum ar fi vârsta, RR este dat pentru o unitate sau creștere de vârstă de 1 an. Riscul relativ pe 5 ani de creștere a vârstei este

$$e^{5\,\text{beta}} \_ e^{5 \times .0598} \_ 135$$

Există o creștere cu 34% a riscului de AVC în viitor pe o vârstă cu 5 ani mai mare la momentul inițial, controlând toate celelalte variabile din model. Pentru a calcula intervalele de încredere pentru acest exemplu, trebuie de asemenea să înmulțiți se cu 5 (precum și să înmulțiți beta cu 5), astfel încât intervalele de încredere de 95% ale RR sunt: [1,18, 1,55] pentru o creștere de 5 ani a vârstei.

Anexa G oferă informații suplimentare despre explorarea unei relații de formă J sau U între o variabilă și rezultat.

4.21 Selectarea variabilelor pentru modelele multivariate

Să presupunem că dorim să determinăm efectul depresiei asupra evenimentelor ulterioare de boli de inimă folosind date dintr-un studiu prospectiv de urmărire. Putem rula modele de riscuri proporționale Cox pentru a obține riscul relativ de depresie pentru efectele bolilor de inimă, dar dorim să controlăm pentru confundatori. În caz contrar, orice asociere pe care o vedem ar putea fi într-adevăr o reflectare a unei alte variabile, cum ar fi fumatul, care este legat de depresie și este, de asemenea, un factor de risc pentru boli de inimă. Cum vom decide ce variabile să introducem în model? Nu există un răspuns unic la această întrebare și experți diferiți au opinii oarecum diferite, deși este în general de acord că ar trebui să fie

incluși factorii de confuzie cunoscuți . Deci am introduce variabile care sunt semnificativ legate de depresie și , de asemenea, de bolile de inimă în rândul celor neexpuși, adică nedepresivi . Nu am include variabile despre care se presupune că din experiența anterioară sunt fie foarte corelate cu depresia (numită coliniară), fie intermediare în calea care leagă depresia și bolile de inimă, cum ar fi un biomarker din sânge legat de bolile de inimă, care este crescută de depresie. Într-un astfel de caz, creșterea biomarkerului sanguin este intermediară între depresie și boli de inimă; poate fi prima manifestare a bolii cardiace. Ideea este că trebuie folosită multă raționament în selectarea variabilelor pentru includere. Obiectivul este de a vedea dacă efectele depresiei care au fost găsite rămân după luarea în considerare a altor factori de risc stabiliți.

O strategie este să începeți prin a obține singur riscul relativ de depresie și apoi adăugați succesiv, unul câte unul, alți potențiali confundatori pentru a vedea dacă modifică riscul relativ de depresie cu 10% sau mai mult (deși acesta este un procent arbitrar). Variabilele care se califică după acest criteriu sunt păstrate în model. De exemplu, în studiul depresiei și deceselor din cauze cardiovasculare, în rândul femeilor în postmenopauză înscrise în Women's Health Initiative, care nu au avut nicio boală cardiovasculară anterioară, riscul relativ asociat cu controlul depresiei pentru vârstă și rasă a fost de 1,58; adăugarea educației și a veniturilor la aceasta a dus la un risc relativ de 1,52. Adăugarea de variabile suplimentare la model (diabet, hipertensiune arterială, fumat, colesterol ridicat care necesită pastile, utilizarea hormonilor, indicele de masă corporală și activitatea fizică) nu a schimbat lucrurile, rezultând un risc relativ de 1,50. Așadar, s-a ajuns la concluzia că depresia este un factor de risc independent pentru moartea cardiovasculară.

Acum, dacă cineva ar fi interesat să dezvolte un model care să prezică riscul (mai degrabă decât unul care să evalueze dacă un anumit factor de risc a contribuit independent la risc, ca în exemplul de mai sus), ar putea alege alte strategii, cum ar fi regresia în trepte. Regresia în trepte poate fi treptat înainte sau înapoi, iar programele de calculator care calculează regresiile vă cer să specificați ce doriți.

Principiul de bază este că, în regresia treptată înainte, introduceți mai întâi singura variabilă care are cea mai mare corelație cu rezultatul dvs., apoi continuați să adăugați variabile pe rând până când adăugați una care nu este semnificativă statistic la un nivel prestabilit, apoi opriți. În regresia

în trepte înapoi, începeți cu toate variabilele potențiale care pot fi explicative și le renunțați pe rând, eliminând-o pe cea mai puțin semnificativă (are cea mai mare valoare p), până când renunțarea la următoarea variabilă ar avea ca rezultat un model mai slab.

Mulți oameni nu le place regresia în trepte, deoarece este oarecum arbitrară; depinde de nivelurile de semnificație pe care le-ați ales să introduceți sau să părăsiți modelul și, de asemenea, o variabilă poate avea un efect destul de diferit dacă este într-un model cu alte variabile care l-ar putea modifica, mai degrabă decât atunci când este doar în model. O altă strategie este să se uite la toate regresiile posibile - adică să se uite la toate cele două modele variabile, apoi la toate cele 3 modele variabile posibile și așa mai departe. Îl selectați pe cel mai bun în funcție de cât de mult din variația variabilei dependente este explicată de model. O discuție excelentă despre selecția variabilelor în modelele epidemiologice este de Sander Greenland [24] și, de asemenea, în textele avansate notate în secțiunea Lecturi sugerate.

4.22 Interacțiuni: modele aditive și multiplicative

O interacțiune între două variabile înseamnă că efectul unei variabile asupra rezultatului de interes este diferit în funcție de nivelul celeilalte variabile, așa cum este descris în Secțiunea 3.26. Interacțiunile pot fi *aditive* , când efectul comun al două variabile este *mai mare decât suma efectelor lor individuale,* sau *multiplicative,* în cazul în care efectul comun al celor două variabile *este mai mare decât produsul efectelor individuale* ale fiecărei variabile .

Modelele de regresie logistică și Cox sunt în mod inerent multiplicative. Când spunem că fumatul prezintă un risc relativ de 2 pentru boala coronariană, de exemplu, ne referim la faptul că fumătorii au șanse de două *ori* mai mari decât nefumătorii de a face boala. Poate dorim să știm dacă există o interacțiune între fumat și hipertensiune arterială. Cu alte cuvinte, vrem să știm dacă fumatul printre hipertensivi are un efect mai mare asupra atacurilor de cord decât ne-am aștepta din cunoașterea riscurilor separate ale fumatului și ale hipertensiunii. Putem testa ipoteza lipsei de interacțiune față de ipoteza alternativă a unei interacțiuni, dar mai întâi trebuie să știm la ce ne-am aștepta de la un model multiplicativ dacă

nu ar exista interacțiune.

Luați în considerare două variabile dihotomice A și B.

Tabelul de mai jos arată modelul riscurilor relative așteptate în cadrul modelului multiplicativ dacă nu există o interacțiune multiplicativă. Grupul de referință este RR $_{nono}$ = 1 (cu alte cuvinte, toate comparațiile noastre sunt la riscul dintre cei care nu au nici A, nici B). Rețineți că RR $_{da\,da}$ = RR $_{da\,nu}$ x RR $_{noyes}$ = 2,0 x 1,5 = 3,0. (RR $_{da\,,\,nu}$ este riscul relativ al lui B în absența lui A și RR $_{nu\,,\,da}$ este riscul relativ al lui A în absența lui B).

Risc relativ (RR)

	A nu	A da
B nr	RR = 1 ^■■^■nr.nr[1 2]	RR = 15 [r " J "-nu, da]
B da	RR $_{da,nu}$ = 3_∘	RR = 3 0 [r ' jx wes.da]

Dacă RR observat $_{da\,da}$ este semnificativ diferit de 3.0, putem respinge ipoteza nulă a lipsei interacțiunii și concluziona că există o interacțiune multiplicativă.

Pentru a testa acest lucru statistic, am include un termen de produs în modelul nostru de regresie, (înmulțind valoarea variabilei B cu valoarea variabilei A pentru fiecare persoană pentru a obține o nouă variabilă care este produsul dintre A și B) și apoi să calculăm următoarea cantitate:

---

eroare standard a /)

Această cantitate pătrat este distribuită aproximativ ca Chi-pătrat cu 2grad de libertate.

Ceea ce testăm cu adevărat este dacă coeficientul / *i* este semnificativ diferit de 0. Dacă este, atunci acest lucru este echivalent cu concluzia că RR $_{da\,da}$ este semnificativ diferit de valoarea noastră așteptată de 3,0

Să presupunem că dorim să vedem ce fel de cifre de incidență ar putea da naștere tabelului de mai sus. Amintiți-vă, incidența este un risc absolut, în timp ce riscul relativ este riscul absolut într-un grup în raport cu cel absolut

(coeficientul termenului de produs în regresia logistică)

risc în grupul de referință. Cele două tabele de mai jos conțin ambele cifre de incidență care ar da naștere tabelului RR de mai sus, așa că puteți vedea că este posibil să existe rate de incidență diferite sau riscuri care au același risc relativ.

Risk or Incidence per 1,000

	A no	A yes
no	$I = 20_{z\check{s}no,no}$	$I = 30$ $_{A no,yes}$
yes	$I = 40$ $_{xyes.no}$ $_{w}$	$I_{yes,yes}$ 60

	A no	A yes
no	$I_{no,no}$ 10	$I_{no,yes}$ 15
yes	$I = 20$ ^{x}vesno	I $= 30$ $_{xyes.yes}$ $_{\sim JXJ}$

Riscul aditiv este mai rar testat, deși unii oameni cred că ar trebui să fie. Se calculează din diferența de riscuri absolute (mai degrabă decât din raportul riscurilor absolute). Sub ipoteza lipsei de interacțiune într-un model aditiv, ne-am aștepta la datele din tabelele de mai jos.

Incidență la 1.000

	A no	A yes
B no	$I = 20_{z\check{s}no,no}$	$I = 30$ $_{A no,yes}$
B yes	$I = 40$ $_{xyes.no}$ $_{w}$	$I_{yes,yes}$ 50

Notă: Incidența $_{da\ da}$ = Incidența de bază + efectul lui A + efectul lui B sau

$I_{yes,yes}$ $I_{no,no}$ + ($I_{no,yes}$ $I_{no,no}$) + ($I_{yes,no}$ - $_{Uno}$) = 20 + 10 + 20 = 50

Diferențele de risc sau riscul atribuibil (AR)

La 1.000

	A nu	A da
B nr	$AR_{nu,\,nu} = 0$	$AR_{nu,\,da} = 10$
B da	$AR_{da,nu} = 20$	$AR_{da,\,da}$ 30

$AR_{da\,da}$ = efectul lui A plus efectul lui B = 10 + 20 = 30

Dacă $AR_{da\,da}$ este suficient de diferit de valoarea așteptată de 30, atunci putem concluziona că există o interacțiune pe scara aditivă.

Tabelul de risc relativ care corespunde tabelului de incidență pentru exemplul dat mai sus al modelului aditiv este:

Risc relativ (RR)

	A nu	A da
B nr	RR = 1 ¹ ^ ʲ ^nu,nu ¹	RR = 15 ¹ ^ ʲ ^nu,da F.-'
B da	$RR_{da,nu} = {}^{2}$ -0	RR = 2 5 F"- ᵉᵘ ''-da,da ^-.-'

Astfel, valoarea așteptată a $RR_{da\,da}$ în ipoteza nulă a interacțiunii fără aditiv este: $RR_{da\,da} = RR_{da\,nu} + RR_{nu\,,\,da} - 1$. Dacă $RR_{da\,da}$ este semnificativ diferit de așteptarea de mai sus, am putea respinge ipoteza nulă a riscului aditiv.

Interacțiunile depind de scară – adică dacă vorbim despre riscuri relative (multiplicative) sau riscuri atribuibile (aditive). Este înțelept să consultați un statistician pentru interpretări adecvate. Surse excelente, mai tehnice, sunt *Epidemiology: Beyond the Basics* de Moyses Szklo și F. Javier Nieto și *Modern Epidemiology* de Rothman și Greenland.

Rezumat:

Efectul de interacțiune al modelului aditiv:

Vedeți dacă valoarea observată a $AR_{da\,da}$ diferă de valoarea așteptată:

AR$_{da,da}$ AR$_{da, nu}$ + AR$_{nu, da}$

sau, RR$_{da, da}$ RR$_{da, nu}$ + RR$_{nu, da}$ -1

Efectul de interacțiune al modelului multiplicativ

Vedem dacă valoarea observată diferă de valoarea așteptată:

RR$_{da, da}$ RR$_{da, nu}$ x RR$_{nu, da}$

Capitolul 5
CELE MAI MULT DESPRE PROIECTARE

Mai degrabă aș profita de șansa mea ca unii trădători să scape de depistare decât să răspândească în străinătate un spirit de suspiciune și neîncredere generală, care acceptă zvonurile și bârfele în locul unei anchete neconsternate și neintimidate.

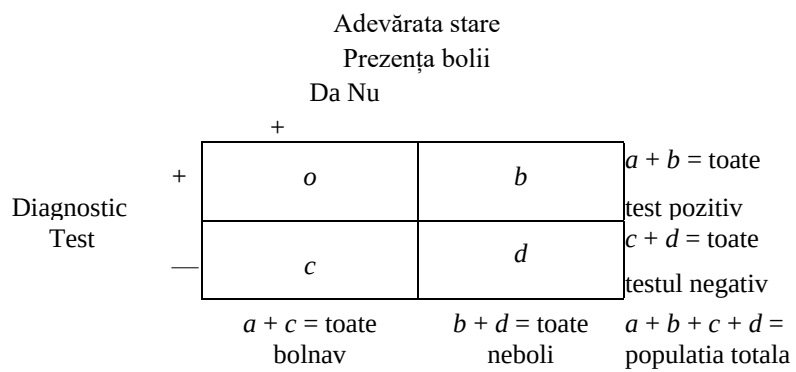
Judecătorul Learned
Hand octombrie 1952

5.1 Sensibilitate, specificitate și concepte înrudite

Problema în utilizarea testelor de screening sau de diagnosticare este de a găsi un compromis adecvat între dorința de a detecta boala la persoanele care o au cu adevărat și dorința de a evita să credeți că ați detectat-o la persoanele care nu o au cu adevărat.

O modalitate importantă de a vizualiza testele de diagnostic și screening este prin analiza sensibilității. Definițiile termenilor și simbolurilor relevanți sunt următoarele:

T+ înseamnă test pozitiv, T— înseamnă test negativ, D+ înseamnă boală, D- înseamnă că nu ai boală.



SENSIBILITATE : proporția de persoane bolnave pe care testul le clasifică drept pozitiv,

$\frac{a}{a + c}$ (probabilitatea testului pozitiv, dată fiind dis-ușurință)

SPECIFICAȚIE : proporția de persoane nebolnave pe care testul le clasifică drept negativ,

$\frac{d}{boala\ b + d}$ (probabilitatea testului negativ, dat nr)

FALS -POZITIVĂ *:* proporția de persoane neboli pe care testul o clasifică (incorect) drept pozitivă,

$\frac{b}{boala\ b + d}$ (probabilitatea testului pozitiv, dat nr)

FALS NEGATIVE *:* proporția de persoane bolnave pe care testul o clasifică (incorect) drept negativ,

$\frac{c}{dis\ a + c}$ (probabilitatea testului negativ dat fiind ușurință)

VALOAREA PREDICTIVA A UNUI TEST POZITIV: proporția de teste pozitive care identifică persoanele bolnave,

$\frac{a}{testul\ a + b}$ (probabilitatea bolii dată pozitiv)

VALOAREA PREDICTIVA A UNUI TEST NEGAT : proporția de teste negative care identifică corect persoanele nebolnave,

$\_\ \underline{d}\ \__{P}D\_|_{T\_}$). (probabilitatea de a nu se da nicio boală
$c + d$ test negativ)

PRECIDEȚIA *TESTULUI* : proporția tuturor testelor care sunt clasificări corecte,

$$\frac{a + d}{a + b + c + d}$$

DE PROBABILITATE *AL TESTULUI POZITIV:* raportul dintre probabilitatea unui test pozitiv, dată fiind boala, și probabilitatea unui test pozitiv, dacă nu există boală,

$$\frac{P(T+ \mid D+)}{P(T+ \mid D-)}$$ test pozitiv, dată boală versus pozitiv , fără boală

sensibilitate	sensitivity
rata fals pozitive	1 - specificity

LIKELIHOOD *RATIO OF A NEGATIVE TEST:*

$\frac{P(T-|D+)}{P(T-|D-)}$ test negativ, având în vedere boala versus test negativ, dat fiind nicio boală

1 -
specificitatea sensibilității

Rețineți și următoarele relații:

(1) Specificitatea + rata fals-pozitive = 1;

$$\frac{d}{b + d} + \frac{b}{b + d} = 1$$

prin urmare, dacă specificitatea unui test este crescută, rata fals-pozitive este scăzută.

(2) Sensibilitate + rata fals-negative = 1;

$$\frac{a}{a + c} + \frac{c}{a + c} = 1$$

prin urmare, dacă sensibilitatea unui test este crescută, rata fals-negative va fi scăzută.

PRETEST *DE BOLI:* Probabilitatea pretest a unei boli este prevalența acesteia. Neștiind nimic despre un individ și în absența unui test de diagnostic, cea mai bună presupunere a probabilității ca pacientul să aibă boala este prevalența bolii.

PROBABILITATEA DE BOLI POSTTEST : După obținerea rezultatelor testului, probabilitatea posttest de boală dacă testul este normal este *c/(c+d*), iar dacă este anormal probabilitatea posttest este *a/(a+b*). Ultima este aceeași cu *VALOAREA PREDICTIVA A UNUI TEST POZITIV.*

Un test de diagnostic bun este unul care vă îmbunătățește estimarea despre starea bolii pacientului față de estimarea pe care ați face-o doar pe baza prevalenței generale a bolii. Cu toate acestea, de interes principal pentru un clinician este *valoarea predictivă a unui test pozitiv (PV+),* care este proporția de persoane care au un test pozitiv și care au cu adevărat boala , *a* /(*a* + *b*), și *valoarea predictivă a unui test negativ (PV-)* , care este proporția de persoane cu un test negativ care nu au boala cu adevărat, *d* /(*c* + *d*).

Sensibilitatea și specificitatea sunt caracteristici ale testului în sine, dar valorile predictive sunt foarte mult influențate de cât de frecventă este boala. De exemplu, pentru un test cu 95% sensibilitate și 95% specificitate folosit pentru a diagnostica o boală care apare doar la 1% dintre oameni (sau 100 din 10.000), am avea următoarele:

		Disease		
		Yes	No	
		+	-	
Test	+	95	495	590
	-	5	9,405	9,410
		100	9,900	10,000

PV+ este 95/590 = .16; asta înseamnă că doar 16% dintre toți oamenii cu rezultate pozitive la test au cu adevărat boala; 84% nu au boala deși testul este pozitiv. PV- cu toate acestea, este de 99,9%, ceea ce înseamnă că dacă

un pacient are un rezultat negativ al testului, poți fi aproape complet sigur că nu are cu adevărat boala. Valoarea practică a unui test de diagnostic depinde de o combinație de sensibilitate, specificitate și prevalență a bolii, toate acestea determinând valorile predictive ale rezultatelor testului.

Dacă prevalența bolii este mare, valoarea predictivă a unui test pozitiv va fi și ea mare, dar un test bun ar trebui să aibă o valoare predictivă mare, chiar dacă prevalența bolii este scăzută. Să aruncăm o privire asupra relației dintre prevalența bolii și sensibilitate, specificitatea și valoarea predictivă a unui test, prezentată în Figura 5.1.

Să luăm, de exemplu, un test care are o sensibilitate de .95 și o specificitate de .99. Aceasta înseamnă că acest test va eticheta corect ca fiind bolnavi 95% dintre persoanele cu boală și va eticheta corect ca nebolnași 99% dintre persoanele fără boală. Să luăm în considerare o boală a cărei prevalență este de 10%, adică 10% din populație are această boală și să ne uităm acum și să vedem care este valoarea predictivă a unui test pozitiv. Remarcăm că este de aproximativ .90, ceea ce înseamnă că 90% dintre indivizii cu un test pozitiv vor avea boala. Putem vedea că valoarea predictivă a unui test pozitiv scade la aproximativ .70 pentru un test care are o sensibilitate de .95 și o specificitate de .95 și putem vedea că mai scade la aproximativ .40 pentru un test care are o sensibilitate de .95 și o specificitate de .85. Cu alte cuvinte, doar 40% dintre indivizii cu un test pozitiv ar avea cu adevărat boala pentru un test care are acea sensibilitate și specificitate specială.

Un lucru pe care îl puteți observa imediat este că, *pentru boala cu prevalență scăzută, valoarea predictivă a unui test pozitiv scade destul de brusc* . Un alt lucru pe care îl puteți observa aproape imediat este că diferența mare de sensibilitate face o mică diferență în valoarea predictivă a unui test pozitiv și că o mică diferență de specificitate face o mare diferență în valoarea predictivă a unui test pozitiv. Aceasta înseamnă că caracteristica unui test de screening descrisă prin specificitate este mai - importantă în determinarea valorii predictive a unui test pozitiv decât sensibilitatea .

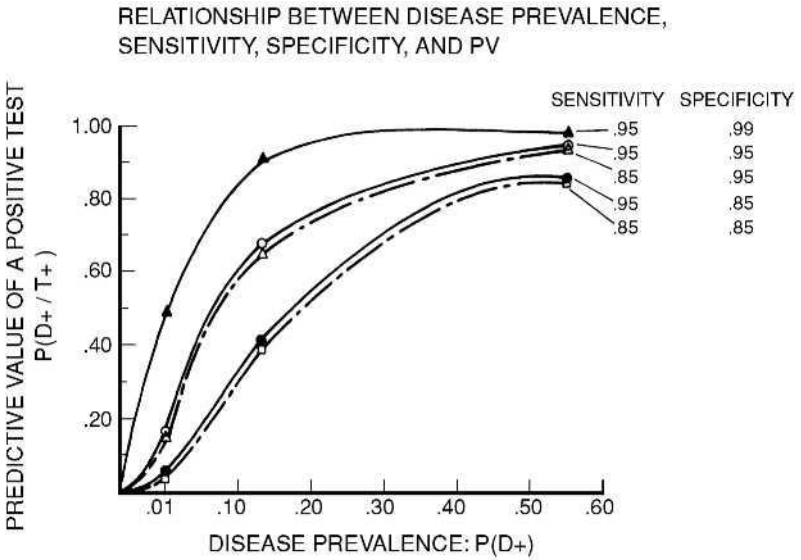


Figura 5.1

Figura 5.2 ne arată o situație a unui test care este practic perfect. Un test care are o sensibilitate de 0,99 și o specificitate de 0,99 este de așa natură încât, la cele mai multe niveluri de prevalență a bolii, probabilitatea bolii, având în vedere un rezultat normal sau negativ al testului, este foarte scăzută. Acesta ar fi un test foarte bun și, cu cât ne putem apropia de acest tip de situație, cu atât testul de diagnosticare este mai bun. Linia diagonală din centru reprezintă un test cu o sensibilitate de .50 și o specificitate de .50 și care, desigur, este un test complet inutil, deoarece puteți observa că la prevalența bolii de .4 probabilitatea ca boala să fie pozitivă la un test este și ea de .4, ceea ce este aceeași cu probabilitatea bolii fără a face niciun test, iar acest nivel de prevalență este valabil. Prin urmare, un astfel de test este complet inutil, în timp ce un test cu sensibilitate și specificitate de .99 este excelent și orice lucru între ele reprezintă o utilitate diferită pentru teste. Acesta este un mod analitic de a analiza procedurile de testare de diagnosticare.

Un exemplu deosebit de relevant al implicațiilor prevalenței asupra

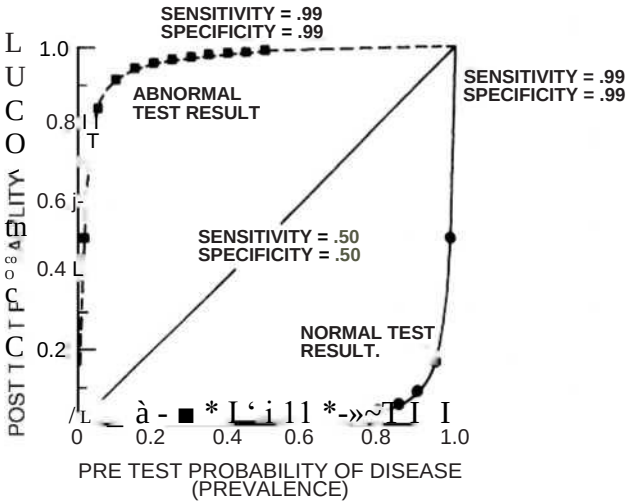


Figure 5.2

valoarea predictivă este cazul screening-ului pentru prezența infecției cu virusul SIDA. Deoarece această boală este în general fatală, incurabilă în prezent, are un stigmat legat de ea și implică costuri mari, nu ar dori să folosim o strategie de screening care etichetează în mod fals oamenii ca fiind pozitivi pentru HIV, virusul SIDA.

Să ne imaginăm că avem un test pentru acest virus care are o sensibilitate de 100% și o specificitate de 99,995%, clar un test foarte bun. Să presupunem că o aplicăm în mod obișnuit tuturor donatoarelor de sânge de sex feminin, la care prevalența este estimată a fi foarte scăzută, 0,01%. În comparație, să presupunem că o aplicăm și bărbaților homosexuali din San Francisco, la care prevalența este estimată la 50%. [22] Pentru fiecare 100.000 de astfel de persoane examinate, am avea valori așa cum se arată în tabelul de pe pagina următoare.

Deși în ambele grupuri ar fi identificați toți cei care au avut cu adevărat boala, printre donatorii de sânge de sex feminin, o treime din totalul persoanelor care au fost testate pozitiv nu ar avea cu adevărat boala; printre homosexuali masculini

Valoarea predictivă pozitivă ca funcție a prevalenței

Caracteristicile testului:

Sensibilitate = 100%; Specificitate = 99,995%; Rata fals pozitive = .005%

A. DONATORI DE SÂNGE FEMINII Prevalența = 0,01%,

stare adevărată

		HIV: +	**HIV: -**	
Rezultatul ecranului	+	10	5	15
		0	99.985	99.985
		10	99.990	100.000

PV+ = 10/15 = .66667

B. HOMOSEXUALI BĂRBAȚI Prevalența = 50%,

Adevărata stare

		HIV: +	**HIV: -**	
Rezultatul ecranului	+	50.000	3	50.003
		0	49.997	49.997
		50.000	50.000	100.000

PV+ = 50.000/50.003 = .99994

sexuali doar 6 din 100.000 de persoane cu un test pozitiv ar fi etichetate fals.

5.2 Cutoff Point și efectele sale asupra sensibilității și specificității

Am discutat despre sensibilitatea și specificitatea ca caracteristice ale unui test de diagnostic; cu toate acestea, ele pot fi modificate prin alegerea

punctului de *tăiere între normal și anormal* . De exemplu, putem dori să diagnosticăm pacienții ca hipertensivi sau normotensivi după tensiunea arterială diastolică. Să spunem că oricine are o presiune diastolică de 90 mm Hg sau mai mult va fi clasificat drept „hipertensiv". Deoarece tensiunea arterială este o caracteristică continuă și variabilă, la orice măsură, un individ de obicei nehipertensiv poate avea o tensiune arterială diastolică de 90 mm Hg sau mai mult și, în mod similar, un individ cu adevărat hipertensiv poate avea o singură măsură mai mică de 90 mm Hg. Cu un punct de limită de 90 mm Hg, vom clasifica unii indivizi nehipertensivi ca hipertensivi și aceștia vor fi fals pozitivi. De asemenea, vom eticheta unii indivizi hipertensivi ca fiind normotensivi și aceștia vor fi fals negativi . Dacă am avea un punct de limită mai strict, să zicem, 105 mm Hg, am clasifica mai puțini nehipertensivi ca hipertensivi, deoarece mai puțini indivizi normotensivi ar avea o citire atât de mare (și ar avea mai puține pozitive false).

Cu toate acestea, am avea mai multe negative false (adică, mai mulți dintre oamenii noștri cu adevărat hipertensivi s-ar putea înregistra ca având o tensiune arterială diastolică mai mică de 105 mm Hg cu orice ocazie). Aceste concepte sunt ilustrate în Figura 5.3.

Există două distribuții ale populației, cea bolnavă și cea nebolnică, și se suprapun pe măsura de interes, indiferent dacă este vorba despre tensiunea arterială, glicemia sau alte valori de laborator. Există foarte puține teste de screening care nu au nicio suprapunere între indivizi normali și bolnavi .

Un obiectiv în luarea deciziei asupra unui punct de limită este de a găsi echilibrul adecvat între fals pozitive și false negative. După cum puteți vedea în Figura 5.3, când punctul de limită este la A, toate valorile din dreapta lui A sunt numite pozitive (se consideră că pacientul are boala). De fapt, totuși , pacientul cu o valoare la dreapta limită A ar putea proveni din populația de persoane nebolnave, deoarece o proporție de oameni care sunt perfect normali pot avea totuși valori mai mari decât cele de deasupra lui A, așa cum se vede în curba normală. Zona din dreapta lui A sub curba fără boală reprezintă fals pozitiv.

Dacă un individ are valoarea est la stânga limitei A, el poate fi un negativ adevărat sau poate fi un negativ fals deoarece o proporție de in-

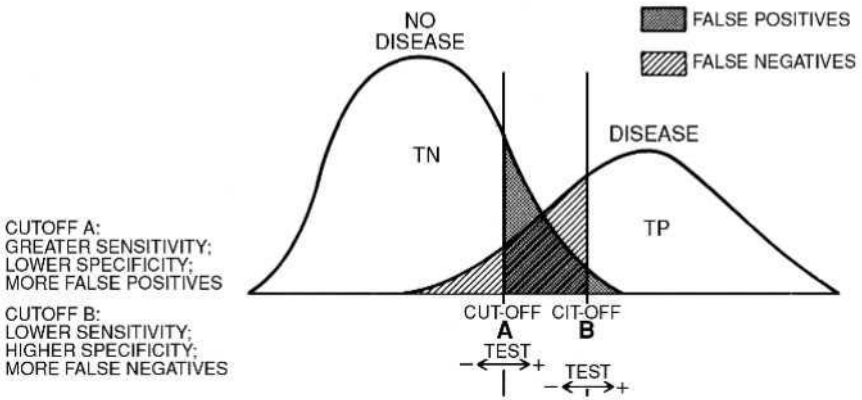


Figura 5.3

indivizii cu boală pot avea totuși valori mai mici decât pragul A. Aria de sub curba „boală" din stânga limitei A reprezintă proporția de fals negative.

Dacă mutăm punctul de tăiere de la A la B, vedem că scădem aria din dreapta limitei, scăzând astfel numărul de fals pozitive, dar crescând numărul de fals negative. În mod corespunzător, cu cutoff-ul A, avem o probabilitate mai mare de a identifica corect bolnavul cu adevărat, adică de a obține mai mult adevărat pozitiv, dând astfel testului cu cut-off A o sensibilitate mai mare. Cu cutoff-ul B, este mai puțin probabil să obținem adevăratul pozitiv (sensibilitate mai scăzută), dar este mai probabil să identificăm corect negativul adevărat (specificitate mai mare).

Astfel, prin deplasarea punctului limită dincolo de ceea ce numim un test pozitiv , putem modifica caracteristicile de sensibilitate și specificitate ale testului. Alegerea limitei, cu excepția cazului în care există un motiv fiziologic special, se poate baza pe luarea în considerare a consecințelor relative ale existenței prea multor fals pozitive sau prea multe fals negative. Într-un test de screening pentru cancer, de exemplu, ar fi de dorit să existe un test de sensibilitate ridicată (și puține fals negative), deoarece eșecul de a detecta precoce această afecțiune este adesea fatală. Într-un test de screening în masă pentru o afecțiune mai puțin gravă sau pentru una în care detectarea precoce nu este critică, poate fi mai de dorit să existe o specificitate ridicată pentru a nu supraîncărca

sistemul de furnizare a asistenței medicale cu prea multe rezultate false pozitive. Luarea în considerare a costurilor poate intra și în alegerea punctului limită.

Relația dintre sensibilitate (capacitatea de a identifica corect indivizii bolnavi) și fracțiile fals pozitive este prezentată în Figura 5.4.

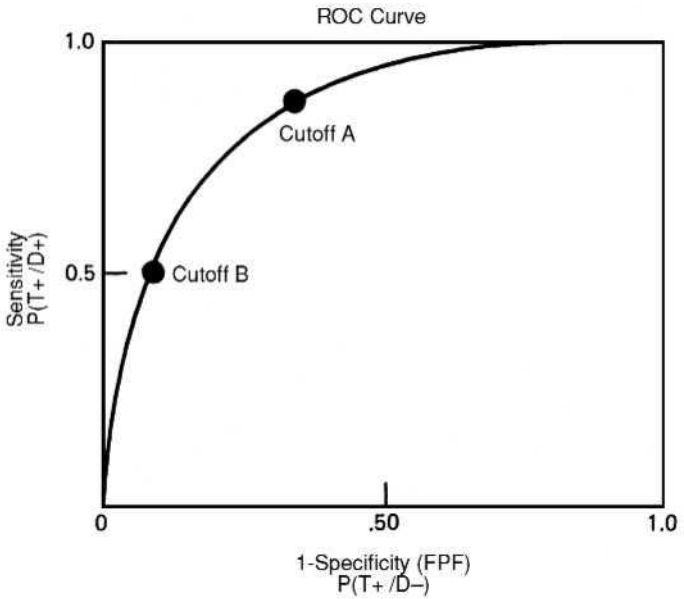


Figura 5.4

Aceasta se numește curba caracteristicii de funcționare a receptorului (ROC) a testului. Adesea, putem selecta punctul de limită între normal și anormal , în funcție de compromisul pe care suntem dispuși să îl facem între sensibilitate și proporția de fals pozitive.

Putem vedea că cu limita A, în timp ce putem detecta un procent mai mare de indivizi cu adevărat bolnavi, vom avea, de asemenea, o proporție mai mare de rezultate fals e-pozitive, în timp ce cu limita B vom avea mai puține fals pozitive, dar vom avea mai puține șanse să detectăm cei cu adevărat bolnavi. Testele de screening ar trebui să aibă trasate curbele ROC corespunzătoare.

Capitolul 6
CELE MAI MULT DESPRE TRESTILE CLINICE

Nu este o sarcină ușoară să-și îndrepte cuiva drumul de la adevăr la adevăr prin erori amenințătoare.

Peter Marc Latham
1789-1875

„Nu l-aș fi văzut dacă nu aș crede!”

Atribuit lui Yogi Berra

Din păcate, uneori, oamenii de știință văd ceea ce cred în loc să trăiască ceea ce văd. Studiile clinice randomizate, controlate au tendința de a evita această părtinire și alte tipuri de părtinire.

Un *studiu clinic randomizat* este un experiment prospectiv pentru a compara una sau mai multe intervenții cu un grup de control pentru a determina eficacitatea intervențiilor. Un studiu clinic poate compara valoarea unui medicament față de un placebo. Un placebo este o substanță inertă care arată ca medicamentul testat. Se poate compara o nouă terapie cu o terapie standard în prezent, chirurgicală cu intervenție medicală, două metode de predare a lecturii, două metode de psihoterapie. Principiile se aplică oricărei situații în care problema cine este expus la care condiție se află sub controlul experimentatorului și metoda de atribuire este prin randomizare.

6.1 Caracteristicile studiilor clinice randomizate

(1) Există un grup de pacienți care sunt desemnați pacienți de studiu. Toate criteriile trebuie stabilite și îndeplinite înainte ca un potențial candidat să poată fi considerat eligibil pentru studiu. Orice excluderi trebuie specificate.
(2) Orice motiv pentru excluderea unui potențial pacient de la participarea la studiu trebuie specificat înainte de începerea studiului. În caz contrar, pot apărea părtiniri neintenționate. De exemplu, să presupunem că comparați operația de bypass coronarian cu utilizarea

unui nou medicament pentru tratamentul bolii coronariene. Să presupunem că vine un pacient care este eligibil pentru studiu și este repartizat la tratamentul chirurgical. Să presupunem că acum descoperi că pacientul are boală de rinichi. Decizi să-l excluzi din studiu pentru că crezi că ar putea să nu supraviețuiască intervenției chirurgicale cu rinichi afectați. Dacă ajungeți să excludeți sistematic toți pacienții mai bolnavi de la tratamentul chirurgical, puteți părăsi rezultatele în favoarea pacienților mai sănătoși, care în orice caz au șanse mai mari de supraviețuire. În acest exemplu, boala renală ar trebui să fie un criteriu de *excludere aplicat pacienților înainte ca aceștia să fie alocați oricărui grup de tratament.*

(3) Odată ce un pacient este eligibil, el sau ea este repartizat aleatoriu în grupul experimental sau de control. Atribuirea aleatorie nu este o atribuire „hap hazard", ci mai degrabă înseamnă că fiecare persoană are șanse egale de a fi un pacient experimental sau de control. De obicei, se realizează prin utilizarea unui tabel de numere aleatorii, descris mai târziu, sau prin numere aleatoare generate de computer.

(4) Studiile clinice pot fi dublu-orb, în care nici medicul curant, nici pacientul nu știu dacă pacientul primește tratament experimental sau placebo; acestea pot fi single-orb, în care medicul curant știe în ce grup se află pacientul, dar pacientul nu știe. Un studiu dublu-orb conține cea mai mică părtinire, dar uneori nu este posibil din motive etice sau practice. De exemplu, medicul poate avea nevoie să cunoască grupul căruia îi aparține pacientul, astfel încât medicația să poată fi ajustată pentru bunăstarea pacientului. Există, de asemenea, studii în care atât pacienții, cât și medicii cunosc grupul de tratament, ca în studiile care compară mastectomia radicală versus lumpectomia pentru tratamentul cancerului de sân. Atunci când rezultatul este mortalitatea, posibila părtinire introdusă este minimă, cu condiția ca criteriile de excludere să fi fost specificate și aplicate înainte ca eligibilitatea să fie determinată în final și ca randomizarea participanților eligibili în grupurile de tratament să fi fost efectuată în mod corespunzător.

(5) În timp ce studiile clinice compară adesea un medicament sau un tratament cu placebo , ele pot compara, de asemenea, două tratamente între ele, sau un tratament și „îngrijire obișnuită". Încercările care compară o intervenție cu „îngrijirea obișnuită", evident, nu pot fi

orbite, de exemplu, comparând o intervenție nutrițională de slăbire cu dieta „obișnuită"; cu toate acestea, evaluarea efectului (măsurarea greutății sau a tensiunii arteriale, sau un efect ipotetic al pierderii în greutate) ar trebui să fie făcută într-un mod orb, evaluatorul neștiind la ce grup a fost repartizat participantul.

(6) Este esențial ca grupul de control să fie cât mai asemănător cu grupul de tratament, astfel încât diferențele de rezultat să poată fi atribuite diferențelor de tratament și nu caracteristicilor diferite ale celor două grupuri. Randomizarea ajută la realizarea acestei comparabilități .

(7) Suntem preocupați aici de studiile de fază III. Noile medicamente trebuie să fie supuse unor studii de fază I și II, care determină toxicitatea și, respectiv, siguranța și eficacitatea. Aceste studii sunt efectuate pe un număr mic de voluntari. Studiile de fază III sunt studii clinice mari, suficient de mari pentru a oferi un răspuns la întrebarea dacă medicamentul testat este mai bun decât placebo sau decât un medicament de comparație.

6.2 Scopurile randomizării

Principiul de bază în proiectarea studiilor clinice sau a oricărei investigații științifice este *evitarea părtinirii sistematice.* Când nu se știe care variabile pot afecta rezultatul unui experiment, cel mai bun mod de a evita prejudecățile sistemice este de a repartiza indivizii în grupuri aleatoriu. *Randomizarea* are scopul de a asigura o distribuție aproximativ egală a variabilelor între diferitele grupuri de indivizi care sunt studiate. De exemplu, dacă studiezi efectul unui medicament antidiabetic și știi că factorii de risc cardiac afectează mortalitatea în rândul diabeticilor, nu ai dori ca toți pacienții din grupul de control să aibă boli de inimă, deoarece asta ar influența în mod clar rezultatele. Prin repartizarea aleatorie a pacienților în grupul de medicament și grupul de control, vă puteți aștepta ca distribuția pacienților cu probleme cardiace să fie comparabilă în cele două grupuri. Deoarece există multe variabile care sunt necunoscute, dar care pot avea o influență asupra rezultatelor, randomizarea este o asigurare împotriva părtinirii necunoscute și neintenționate. Desigur, atunci când se ocupă de variabile cunoscute a fi relevante, se pot lua în considerare acestea prin *stratificarea și apoi randomizarea în cadrul straturilor* . De exemplu,

vârsta este o variabilă relevantă pentru rezultatul diabetului. Pentru a stratifica în funcție de vârstă, puteți selecta patru grupe de vârstă pentru studiu: 35-44, 45-54, 55-64, 65 plus. Fiecare grup este considerat un strat. Când un pacient intră în studiul clinic, stratul său de vârstă este mai întâi determinat și apoi este repartizat aleatoriu fie în grupul experimental, fie în grupul de control. Sexul este o altă variabilă care este adesea tratată prin stratificare.

Un alt scop al randomizării are de-a face cu faptul că tehnicile statistice utilizate pentru compararea rezultatelor între grupurile de pacienți studiate sunt valabile în baza anumitor ipoteze care decurg din randomizare. Motivele matematice ale acestui lucru pot fi găsite în textele mai avansate enumerate în lecturi sugerate.

Trebuie amintit că uneori randomizarea nu reușește să rezulte în grupuri comparabile din cauza întâmplării. Acest lucru poate prezenta o problemă majoră în interpretarea rezultatelor, deoarece diferențele de rezultat pot reflecta diferențe în compoziția grupurilor în ceea ce privește caracteristicile de bază, mai degrabă decât efectul intervenției. Sunt disponibile metode statistice pentru a ajusta caracteristicile de referință despre care se știe că sunt legate de rezultat. Unele dintre aceste metode sunt regresia logistică, modelele de riscuri proporționale Cox și analizele de regresie multiplă.

6.3 Cum se efectuează o atribuire aleatorie

Alocarea aleatorie într-un grup experimental sau într-un grup de control înseamnă că fiecare individ eligibil are șanse egale de a fi în fiecare dintre cele două grupuri. Acest lucru este adesea realizat prin utilizarea tabelelor cu numere aleatorii. De exemplu, un extras dintr-un astfel de tabel este prezentat mai jos:

48461	70436	04282
76537	59584	69173

Utilizarea sa ar putea fi după cum urmează. Toate persoanele cu număr par sunt alocate grupului de tratament și toate persoanele cu număr impar sunt alocate grupului de control. Prima persoană care intră în studiu primește primul număr din listă, următoarea persoană primește următorul număr și

așa mai departe. Astfel, primei persoane i se atribuie numărul 48461, care este un număr impar și atribuie pacientul grupului de control. Următoarea persoană primește 76537; acesta este și un număr impar, așa că și el aparține grupului de control. Următoarele trei persoane care au intrat în studiu au toate numere pare și sunt în grupul experimental. Pe termen lung, va exista un număr egal de pacienți în fiecare dintre cele două grupuri.

6.4 Teste cu două cozi versus test cu o singură coadă

Un studiu clinic este conceput pentru a testa o anumită ipoteză. Se vede adesea această frază în articolele de cercetare: „Semnificativ la nivel .05, test cu două cozi". Amintiți-vă că într-o secțiune anterioară am discutat conceptul de „ipoteză nulă", care afirmă că nu există nicio diferență între două grupuri pe o măsură de interes. Am spus că pentru a testa această - ipoteză vom aduna date astfel încât să putem decide dacă ar trebui să respingem ipoteza lipsei de diferență în favoarea unei ipoteze alternative. *Un test cu două cozi versus un test cu o singură coadă se referă la ipoteza alternativă propusă.* De exemplu, să presupunem că sunteți interesat să comparați nivelul mediu de colesterol al unui grup tratat cu un medicament pentru scăderea colesterolului cu media unui grup de control căruia i sa administrat un placebo. Veți colecta datele corespunzătoare dintr-un studiu bine conceput și ați configura ipoteza nulă ca

H_o: colesterol mediu în grupul tratat = colesterol mediu în control grup.

Puteți alege ca ipoteză alternativă

H_A: colesterolul mediu în grupul tratat este *mai mare decât* media la martori.

În această împrejurare, ați respinge ipoteza nulă în favoarea ipotezei alternative dacă media observată pentru femei ar fi suficient de *mare* decât media observată pentru bărbați, pentru a vă conduce la concluzia că o diferență atât de mare în acea direcție nu este probabil să se fi produs doar întâmplător. Acesta, deci, ar fi un test cu o singură coadă al ipotezei nule.

Dacă, totuși, ipoteza dvs. alternativă a fost că nivelul mediu de colesterol pentru femei este *diferit* de nivelul mediu de colesterol pentru bărbați, atunci ați respinge ipoteza nulă în favoarea alternativei fie *dacă* media pentru femei a fost *suficient de mai mare* decât media pentru bărbați, *fie* dacă media pentru femei a fost *suficient de mai mică* decât media pentru bărbați. Direcția diferenței nu este specificată. În cercetarea medicală este mai obișnuit să folosim un test de semnificație cu două cozi, deoarece adesea nu știm în ce direcție se poate dovedi o diferență, chiar dacă putem crede că știm înainte de a începe experimentul. În orice caz, este important să raportăm dacă folosim un test cu o coadă sau un test cu două cozi.

6.5 Studiul clinic ca „Standard de aur”

Uneori, dovezile din studiile observaționale pot duce la concluzii înșelătoare cu privire la eficacitatea sau siguranța unui tratament, doar pentru a fi răsturnate de dovezile din studiile clinice, cu implicații enorme pentru sănătatea publică. Studiul clinic al inițiativei pentru sănătatea femeilor (WHI) privind terapia hormonală este un exemplu dramatic în acest sens. [21] Estrogenul a fost aprobat de FDA pentru ameliorarea simptomelor post-menopauză în 1942, comercializat agresiv la mijlocul anilor 1960, iar după 1980, în general, combinat cu progesteron pentru femeile cu uter, deoarece s-a constatat că progestinul compensa riscurile de estrogen pentru cancerul uterin. Între timp, multe studii prospective mari de urmărire au arătat aproape uniform că estrogenul reduce bolile de inimă cu 30-50%. În 1993, WHI a organizat un studiu clinic amplu pentru a răspunde cu adevărat la întrebarea riscurilor și beneficiilor pe termen lung ale terapiei hormonale. O parte a fost studiul numai estrogenului pentru femeile care au suferit o histerectomie și, prin urmare, nu au avut nevoie de progesteron pentru a-și proteja uterul, iar o altă parte a fost estrogen plus progesteron (E+P) pentru femeile cu uterul intact.

Studiul E+P a fost un studiu clinic randomizat, dublu orb, controlat cu placebo, menit să se desfășoare pe o perioadă medie de 8,5 ani. Acesta a inclus 16.608 femei cu vârste cuprinse între 50-79 de ani; o dimensiune atât de mare a eșantionului a fost considerată necesară

sary pentru a obține o putere adecvată. Studiul a fost oprit în 2002, cu trei ani înainte de finalizarea sa planificată, deoarece Consiliul de monitorizare

a datelor și siguranței sau DSMB (așa cum este descris în capitolul 9) a descoperit că estrogenul plus progestativ au cauzat un exces de cancer de sân și, în mod surprinzător, a existat un exces semnificativ și complet neașteptat de atacuri de cord în grupul E+P în comparație cu placebo! Rezultatele finale, raportate în lucrările ulterioare, au arătat că efectele adverse (o creștere cu 24% a cancerului de sân invaziv, creșterea cu 31% a accidentelor vasculare cerebrale, 29% a creșterii bolilor coronariene și o creștere de peste două ori a emboliei pulmonare și a demenței) au compensat beneficiile (o scădere cu 37% a cancerului de sân invaziv, scădere cu 37% a numărului de fracturi de șold, luate împreună, astfel încât, luate împreună, scăderea numărului de cancere și fracturi de șold34%) evenimentele dăunătoare pe an a fost substanțială. Deoarece au existat 6 milioane de femei care au luat acest preparat numai în Statele Unite și alte milioane la nivel global, aceste rezultate au implicații importante pentru alte femei decât cele din studiu în sine.

De ce rezultate atât de diferite dintr-un studiu clinic decât din studii longitudinale observaționale? Cea mai probabilă explicație este prejudecata de selecție. Femeile care luau hormoni și apoi urmau să-și observe ratele bolilor de inimă au fost în practic toate studiile observaționale , mai sănătoase, mai slabe, mai active, mai educate, mai puțin supraponderale decât omologii lor fără hormoni, iar stilul lor de viață mai sănătos și starea de sănătate de bază mai bună decât hormonii în sine, a fost ceea ce a explicat ratele lor mai scăzute de boli de inimă.

La întrebarea se răspunde acum folosind „standardul de aur”, studiul clinic: estrogenul plus progestin nu protejează împotriva bolilor de inimă și, de fapt, crește riscul. După cum sa menționat anterior, impactul acestei cercetări este mare, deoarece atât de multe milioane de femei foloseau preparatul testat.

6.6 Regresia spre medie

Când selectați dintr-o populație acei indivizi care au hipertensiune arterială și, ulterior, măsurați din nou tensiunea arterială, media celor de-a doua măsurători va tinde să fie mai mică decât media primelor măsurători și va fi mai apropiată de media populației inițiale din care au fost extrași acești indivizi. Dacă între prima și a doua măsurători ați instituit un tratament, este posibil să atribuiți incorect scăderea tensiunii arteriale medii în grup

efectelor tratamentului, în timp ce o parte din această scădere se poate datora fenomenului numit *regresie spre medie* . (Acesta este unul dintre motivele pentru care un grup de control placebo este cel mai important pentru compararea efectelor tratamentului de mai sus și dincolo de cele cauzate de regresia la medie.) Regresia la medie apare atunci când selectați un grup, deoarece indivizii au valori care se situează peste un anumit nivel de criteriu, ca în screening. Se datorează variabilității erorii de măsurare. Luați în considerare tensiunea arterială.

Valoarea observată a tensiunii arteriale este valoarea adevărată a persoanei plus o cantitate necunoscută de eroare. Presupunerea este că tensiunea arterială măsurată a oamenilor este în mod normal distribuită în jurul valorii medii a valorii lor adevărate, dar necunoscute, a tensiunii arteriale. Să presupunem că vom include oameni în studiul nostru doar dacă tensiunea arterială este de 160 sau mai mult. Acum să presupunem că tensiunea arterială sistolică adevărată a cuiva este 150, dar o măsurăm 160. Selectăm acea persoană pentru grupul nostru de studiu doar pentru că valoarea sa măsurată este mare. Cu toate acestea, data viitoare când îi vom măsura tensiunea arterială, probabil că el va fi mai aproape de valoarea sa reală de 150 decât prima dată. (Dacă ar fi fost aproape de valoarea sa adevărată de 150 prima dată, nu l-am fi selectat niciodată pentru studiul nostru, pentru că ar fi fost sub punctul nostru de limită. Așa că trebuie să fi avut o eroare mare la prima măsurătoare.) Deoarece aceste erori sunt distribuite în mod normal în jurul valorii medii reale de 150, data viitoare avem mai multe șanse să obținem o eroare mai mică și, prin urmare, o presiune arterială mai mică și, prin urmare, o presiune mai mică decât 160.

Să presupunem acum că selectăm un întreg subgrup de oameni care au valori ridicate. Mediile celor doua măsurători ale acestor persoane selectate vor tinde să fie mai mici decât media primelor lor măsurători și mai apropiate de media întregului grup din care le-am selectat. Ideea este că persoanele care au cele mai mari valori prima dată nu au întotdeauna cele mai mari valori a doua oară, deoarece corelația dintre prima și a doua măsurătoare nu este perfectă . În mod similar, dacă selectăm un grup de persoane din cauza valorilor scăzute ale unei anumite caracteristici, media celor doua măsurători la acești oameni va fi mai mare decât media primelor măsurători și, din nou, mai aproape de media întregului grup.

O altă explicație a acestui fenomen poate fi ilustrată prin următorul exemplu de aruncare a unui zar. Imaginează-ți că arunci un zar de 360 de

ori. Ori de câte ori zarul ajunge pe un cinci sau un șase, vei arunca din nou zarul. Suntem interesați de trei medii diferite: (1) media primelor 360 de aruncări, (2) media aruncărilor care vor avea ca rezultat aruncarea noastră din nou și (3) media celor de-a doua aruncări. Rezultatele noastre sunt prezentate în tabelul de pe pagina următoare.

Deși la prima aruncare media celor 360 de ori este 3,5, alegem doar cele două numere cele mai mari, iar media *lor* este 5,5. Aceste 120 de ori când zarul a aterizat pe 5 sau 6 ne vor face să aruncăm din nou, dar la a doua aruncare rezultatul poate varia liber între 1 și 6. Prin urmare, media celei de-a doua aruncări trebuie să fie mai mică decât media grupului din care am selectat la prima aruncare în special pentru că avea valorile mari.

Prima aruncare		A doua aruncare	
Rezultat	Numărul de ori rezultatul este obținut	Rezultat	Numărul de ori rezultatul este Obținut
1	60		
2	60		
3	60		
4	60		
5	***60***	1	20
6	***60***	2	20
		3	20
		4	20
		5	20
Medi de 360 aruncări = 3,5		6	20
Media celor 120 de aruncări care au aterizat 5 sau 6 = 5,5		În medie	A 2-a aruncare = 3,5

6.7 Analiza intentiei de tratare

Datele din studiile clinice în general ar trebui analizate prin compararea grupurilor așa cum au fost inițial randomizate, și nu prin compararea cu grupul de control placebo doar a celor din grupul de droguri care au luat efectiv medicamentul. Persoanele alocate grupului de droguri active ar

trebui incluse în acel grup pentru analiză, chiar dacă nu au luat niciodată medicamentul. Acest lucru poate suna ciudat, deoarece cum se poate evalua eficacitatea unui medicament dacă pacientul nu îl ia? Dar tocmai motivul pentru care oamenii ar putea să nu se conformeze regimului de droguri poate avea de-a face cu efectele adverse ale medicamentului, astfel încât, dacă îi selectăm numai pe cei care se conformează, avem un grup diferit de cel randomizat și putem avea o imagine părtinitoare a efectelor medicamentului.

Un alt aspect este că poate exista o anumită calitate a complianților în general care afectează rezultatul. Un exemplu faimos de concluzii înșelătoare care ar putea apărea din neefectuarea unei analize intenționate de tratare provine din Proiectul de medicamente coronariene. [26] Acest studiu randomizat, dublu -orb a comparat medicamentul clofibrat cu placebo pentru reducerea colesterolului. Variabila rezultatului, care a fost mortalitatea pe cinci ani, a fost foarte asemănătoare în ambele grupuri, 18% în grupul cu medicamente și 19% în grupul placebo. S-a dovedit, totuși, că doar aproximativ două treimi dintre pacienții care ar fi trebuit să ia clofibrat au fost de fapt complianți și și-au luat medicamentele. Acești oameni au avut o rată a mortalității de 15%, semnificativ mai mică decât mortalitatea de 19% în grupul placebo. Cu toate acestea, analize suplimentare au arătat că dintre cei alocați grupului placebo, o treime nu și-au luat nici pastilele placebo. Cele două treimi din grupul placebo care s-au conformat au avut o mortalitate de 15%, la fel ca și cei care s-au conformat cu medicamentul clofibrat! Persoanele care nu au respectat atât grupul cu medicament, cât și grupul placebo au avut o mortalitate mai mare (25% pentru clofibrat și 28% pentru placebo). Poate fi de dorit în anumite circumstanțe să se analizeze efectul unui medicament la cei care îl iau efectiv. În acest caz, comparația dintre cei care respectă medicamentele ar trebui să fie mai degrabă cu cei care respectă placebo decât cu grupul placebo în ansamblu.

Includerea neconformităților în analiză diluează efectele, așa că ar trebui depuse toate eforturile pentru a minimiza neconformitatea. În unele studii, capacitatea evaluată de conformare este un criteriu de înscriere și se face o evaluare a fiecărui pacient ca parte a stabilirii eligibilității sale pentru a stabili dacă acest pacient este probabil să adere la regim. Cei care nu ar putea face acest lucru sunt excluși înainte de randomizare. Cu toate acestea, dacă întrebarea este cât de acceptabil este tratamentul pentru pacient, pe lângă eficacitatea acestuia, atunci baza pentru includere poate fi populația

generală care ar putea beneficia de medicament, inclusiv cei care nu respectă.

În Inițiativa pentru sănătatea femeilor, analiza principală a fost intenția de tratare. Cu toate acestea, o analiză secundară ajustată pentru conformitate (denumită mai frecvent aderență). În această analiză, istoricul evenimentului al participantei a fost cenzurat la șase luni după ce ea fie a încetat să mai ia pastilele de studiu, fie a luat mai puțin de 80% dintre pastilele de studiu. În grupul placebo, istoricul evenimentului a fost cenzurat la șase luni după ce participantul a început să ia hormoni (unii participanți din grupul placebo au încetat să ia pastilele de studiu, dar li s-au prescris hormoni de către medici și au început să le ia singuri). Astfel, această analiză secundară a comparat, practic, cele două grupuri „așa cum au fost tratate", mai degrabă decât cele atribuite unui anumit tratament. În analiza cu intenție de tratament, raportul de risc pentru boala coronariană a fost de 1,24, în timp ce în analiza „ajustată cu aderența" a fost de 1,50. Astfel, constatările din analiza intenției de tratare au fost confirmate și consolidate în analizele ajustate în funcție de aderență.

6.8 Cât de mare ar trebui să fie studiul clinic?

Un studiu clinic ar trebui să fie suficient de mare, adică să aibă o dimensiune a eșantionului suficient de mare, pentru a avea o probabilitate mare de a detecta o diferență *reală* între cele două grupuri. Dacă faceți o mică încercare și nu găsiți nicio diferență semnificativă, nu ați obținut informații noi; este posibil să nu fi găsit o diferență pur și simplu pentru că nu ați avut suficienți oameni în studiu. Nu puteți afirma că nu există nicio diferență între tratamente. Dacă aveți un proces mare și nu găsiți nicio diferență semnificativă, atunci puteți spune cu mai multă certitudine că tratamentele nu sunt într-adevăr diferite.

Să presupunem că găsiți o diferență semnificativă într-un studiu mic cu $p < .05$ (nivel de semnificație). Aceasta înseamnă că rezultatul pe care l-ați obținut este probabil să fi apărut pur întâmplător de mai puțin de 5 ori din 100 (dacă nu ar exista într-adevăr nicio diferență). Este de încredere la fel de mult ca aceeași valoare *p* dintr-un proces mare? Există mai multe școli de gândire despre asta.

Valoarea *p* este un indice al puterii dovezilor cu privire la respingerea unei ipoteze nule. Unii oameni cred că o valoare *p* este o valoare *p* și are aceeași pondere, indiferent dacă provine dintr-un studiu mare sau mic. Alții cred că, dacă obțineți un rezultat semnificativ într-un studiu mic, înseamnă că efectul (sau diferența dintre două medii ale populației) trebuie să fie suficient de mare încât să puteți detecta chiar și cu eșantioanele dumneavoastră mici și, prin urmare, este o diferență semnificativă . Este adevărat că, dacă dimensiunea eșantionului este suficient de mare, putem găsi semnificație statistică dacă diferența reală dintre medii este foarte, foarte mică și practic irelevantă. Prin urmare, găsirea unei diferențe semnificative într-un studiu mic înseamnă că efectul a fost relativ mare.

Alții spun că, în practică, totuși, puteți avea mai puțină încredere că tratamentele diferă într-adevăr pentru o anumită valoare *p* într-un studiu mic decât dacă ați fi obținut aceeași valoare *p* la testarea acestor două tratamente într-un studiu mare. [27] Acest aparent paradox poate apărea în situațiile în care se desfășoară mult mai multe studii mici în întreaga lume care studiază aceeași problemă decât studii mari, cum ar fi în terapia cancerului. Unele dintre aceste încercări, doar întâmplător, se vor dovedi a avea rezultate semnificative care pot induce în eroare.

Să presupunem că există 1.000 de teste mici de terapie cu medicamente anticancer . Din întâmplare, 5% dintre acestea vor fi semnificative chiar dacă terapiile nu au niciun efect, sau 50 de rezultate semnificative. Deoarece acestea sunt doar întâmplătoare, înseamnă că suntem incorecți să declarăm efectele medicamentelor anticanceroase în aceste studii (am comis erori de tip I). Să presupunem, în plus, că există doar 100 de studii mari care studiază aceeași problemă. Dintre acestea, 5%, sau cinci astfel de studii, vor declara că există o diferență, în mod incorect. Deci, dacă combinăm toate încercările care arată diferențe semnificative *în mod incorect* , avem 55 de astfel de valori *p semnificative, dar înșelătoare.* Dintre acestea, 50 sau 91% provin din încercări mici și 5 din cele 55 incorecte (sau 9%) provin din încercările mari. Următoarele puncte sunt importante:

(1) Există o distincție între *semnificația statistică* și *semnificația clinică.* Este posibil ca un rezultat să nu fi apărut întâmplător, adică să reflecte o diferență adevărată, dar să fie atât de mic încât să nu-l facă să nu aibă importanță practică.

(2) Cel mai bine este să raportați probabilitatea reală de a obține

rezultatul doar întâmplător sub ipoteza nulă, adică *valoarea reală p* , mai degrabă decât să spuneți că este semnificativ sau nu. Valoarea *p* pentru ceea ce numim în mod obișnuit „semnificație" este arbitrară. Prin personalizare, sa considerat o valoare *p* de .05 sau mai puțin. Dar punctul de limită .05 nu este sacru. Cititorul ar trebui să decidă ce putere va pune în dovezile furnizate de studiu, iar cititorul trebuie să aibă informațiile necesare pentru a lua acea decizie. Informațiile trebuie să includă proiectarea studiului, selecția eșantionului, dimensiunile eșantionului, abaterile standard și valorile reale *ale p* .

În concluzie:

(1) Găsirea *unei diferențe semnificative* dintr-un mic proces nu ne spune nimic.
(2) Găsirea *unei diferențe semnificative* într-un studiu mare este o constatare reală și ne spune că este posibil ca tratamentele să fie echivalente.
(3) Găsirea unei *diferențe semnificative* într-un studiu mic poate fi sau nu replicabilă.
(4) Găsirea unei *diferențe semnificative* într-un proces mare este de încredere ca dezvăluirea unei adevărate diferențe.

6.9 Ce este implicat în calculul dimensiunii eșantionului?

a. Dimensiunea efectului

Să spunem că 15% dintre victimele unui anumit tip de atac de cord mor dacă li se administrează medicamentul A și 16% mor dacă li se administrează medicamentul B. Această diferență de 1% înseamnă că medicamentul A este mai bun? Majoritatea oamenilor ar spune că aceasta este o diferență prea mică, chiar dacă nu apare întâmplător, pentru a avea vreo importanță clinică. Să presupunem că diferența dintre cele două medicamente este de 5%. Am spune acum că medicamentul A este mai bun? Asta ar depinde de cât de mare diferența am considerat-o importantă. Mărimea diferenței pe care vrem să o detectăm se numește *dimensiunea efectului* .

Pentru a calcula dimensiunea eșantionului, trebuie să cunoașteți

dimensiunea minimă a diferenței dintre două tratamente pe care ați fi dispus să *nu* le detectați . Să presupunem, de exemplu, că în grupul dumneavoastră de control 30% dintre pacienții fără tratament își revin. Convingerea dumneavoastră este că cu tratamentul în grupul experimental 40% se vor recupera. Crezi că această diferență în rata de recuperare este importantă din punct de vedere clinic și vrei să fii sigur că poți detecta o diferență cel puțin la fel de mare ca diferența dintre 30% și 40%. Aceasta înseamnă că, dacă rata de recuperare a grupului de tratament ar fi de 35%, ați fi dispus să nu găsiți acel efect mic. Cu toate acestea, dacă rata de tratament a fost de 40% sau mai mult, ați dori să fiți destul de sigur că o găsiți. Cât de sigur ai vrea să fii? Problema „cât de sigur" are de-a face cu „puterea" testului statistic.

b. Putere

Puterea statistică înseamnă *probabilitatea* de a găsi un efect real (de mărimea pe care o considerați importantă clinic). Relațiile dintre putere, nivelul de semnificație și eroarea de tip I și tip II sunt rezumate mai jos:

Nivel de semnificație = probabilitatea unei erori de tip I = probabilitatea de a găsi un efect atunci când într-adevăr nu există. Acest lucru este cunoscut și sub numele de al pha sau *a.*

Probabilitatea erorii de tip II = probabilitatea de a nu găsi un efect atunci când există într-adevăr unul. Acest lucru este cunoscut și sub numele de beta sau *[5.*

Puterea = 1 - probabilitatea erorii de tip II = probabilitatea de a găsi un efect atunci când există într-adevăr unul. Acesta este, de asemenea, cunoscut sub numele de *1 - beta.*

c. Dimensiunea eșantionului

Pentru a calcula dimensiunea eșantionului, trebuie să specificați dimensiunea efectului, nivelul de semnificație și puterea dorită . Dacă alegeți un nivel de semnificație de .05 și o putere de .80, atunci probabilitatea dvs. de eroare de tip II este 1 - putere sau .20. Aceasta înseamnă că considerați că o eroare de tip I este de patru ori mai gravă decât

o eroare de tip II (.20/.05 = 4) sau că vă este de patru ori mai frică să găsiți ceva care nu există decât să nu găsiți ceva care este. Când calculați dimensiunea eșantionului, există întotdeauna un compromis. Dacă doriți să reduceți probabilitatea de a face o eroare de tip I, atunci pentru o anumită dimensiune a eșantionului și dimensiunea efectului, veți crește probabilitatea de a face o eroare de tip II. Puteți menține la un nivel scăzut ambele tipuri de erori prin creșterea dimensiunii eșantionului. Partea de sus a tabelului de pe pagina următoare arată dimensiunile eșantionului necesare pentru a compara două grupuri cu un test între două proporții în ipoteze diferite .

Al doilea rând al tabelului arată că dacă doriți să puteți detecta o diferență în rata de răspuns de la 30% în grupul de control la 50% sau mai mult în grupul de tratament cu o probabilitate (putere) de 0,80, veți avea nevoie de 73 de persoane în fiecare dintre cele două grupuri. Dacă, totuși, vrei să fii destul de sigur că găsești o diferență la fel de mică precum cea dintre 30% și 40%, atunci trebuie să ai 280 de persoane în fiecare grup.

Dacă doriți să fiți mai siguri că găsiți diferența, să spunem 90% sigur în loc de 80% sigur, atunci veți avea nevoie de 388 de persoane în fiecare grup (în loc de 280 pentru putere .80). Dacă doriți să aveți un nivel de semnificație mai strict de .01, veți avea nevoie de 118 persoane în fiecare grup (comparativ cu cele 73 necesare pentru nivelul de semnificație .05) pentru a putea detecta diferența dintre 30% și 50%; veți avea nevoie de 455 de persoane (comparativ cu 280 pentru nivelul .05) pentru a detecta o diferență de la 30% la 40% rata de răspuns.

Partea de jos a tabelului de pe pagina următoare arată impactul asupra dimensiunii eșantionului al unui test de semnificație cu o singură coadă față de un test cu două cozi. Amintiți-vă că un test cu două cozi postulează că rata de răspuns în grupul de tratament poate fi *fie mai mare, fie mai mică* decât rata de răspuns în grupul de control, în timp ce un test cu o singură coadă specifică direcția diferenței ipotetice. *Un test cu două cozi necesită o dimensiune mai mare a eșantionului* , dar acesta este cel mai frecvent utilizat.

Exemple de dimensiuni de eșantion				
Semnificație Nivel (1 coadă)	Presupunem:	Dimensiunea efectului	Putere	Dimensiunea eșantionului
	Rata de răspuns a grupului de control =	Detectează creșterea grupului de tratament cel puțin pentru:	Cu Probabilitatea de:	*n* Necesar în fiecare grup
.05	30%	40%	.80	280
	30%	50%	.80	73
	30%	40%	.90	388
	30%	50%	.90	101
.01	30%	40%	.80	455
	30%	50%	.80	118
	30%	40%	.90	590
	30%	40%	.90	153

Exemple de dimensiuni de eșantion				
Nivel de semnificație = .05	Presupunem:	Dimensiunea efectului	Putere	Dimensiunea eșantionului
	Rata de răspuns a grupului de control =	Detecta Creșterea grupului de tratament cel puțin la:	Cu Probabilitate de:	*n* Necesar în Fiecare Grup
1-coada	30%	40%	.80	280
2-cozi	30%	40%	.80	356
1-coada	30%	50%	.80	73
2-cozi	30%	50%	.80	92

d. Câteva considerații suplimentare

Pentru o dimensiune fixă a eșantionului și o anumită dimensiune a efectului sau diferență pe care doriți să o detectați, puterea maximă apare atunci când rata evenimentului este de aproximativ 50%. Deci, pentru a maximiza puterea, poate fi uneori înțelept să selectați un grup pentru studiu care este probabil să aibă evenimentele de interes. De exemplu, dacă doriți să studiați efectele unui medicament beta-blocant asupra prevenirii atacurilor de cord, puteți obține „mai multă putere pentru bani" studiind persoanele care au avut deja un atac de cord, mai degrabă decât persoanele sănătoase, deoarece primele sunt mai susceptibile de a avea un alt eveniment (atac de cord). Desigur, atunci s-ar putea să vă uitați la o altă întrebare, efectul beta- -blocantelor asupra supraviețuitorilor atacului de cord, (care ar fi un traseu de prevenire secundară), mai degrabă decât efectul beta-blocantelor în prevenirea primului atac de cord. (un studiu de prevenire primară). Uneori, un studiu de prevenire primară oferă un răspuns diferit decât un studiu de prevenire secundară. Este posibil să puteți interveni pentru a preveni îmbolnăvirea în rândul persoanelor care nu suferă încă de boală, dar intervenția dvs. poate avea un efect redus asupra unei persoane care a dezvoltat deja boala. În mod clar este nevoie de judecată.

6.10 Cum se calculează dimensiunea eșantionului pentru diferența dintre două proporții

Trebuie să specificați care credeți că este probabil să fie proporția de evenimente în fiecare dintre cele două grupuri comparate. Un eveniment poate fi un răspuns, o moarte, o recuperare – dar trebuie să fie o variabilă dihotomică. Specificarea ratelor evenimentelor din cele două grupuri reflectă dimensiunea diferenței pe care ați dori să o puteți detecta.

Specifica:

p_1 = rata în grupa 1; $q_1 = 1 - p_1$; alfa = nivel de semnificație

p_2 = rata în grupa 2; $q_2 = 1 - p_2$; putere

$$n = \frac{(P1Q1) + (P2\text{^}2)}{(P_2 - A)} 2\text{—} \times f\,(alfa, putere).$$

Valorile lui *f* (alfa, putere) pentru un test cu două cozi pot fi obținute din tabelul de mai jos.

Values of *f* (alpha, power)

		.95	.90	.80	.50
Alpha Significance Level .01	.10	10.8	8.6	6.2	2.7
	.05	13.0	10.5	7.9	3.8
	.01	17.8	14.9	11.7	6.6

Notă: *n* este aproximativ invers proporțional cu $(p - p_1)^2$.

Exemplu. Să presupunem că doriți să găsiți dimensiunea eșantionului pentru a detecta o diferență de la 30% la 40% între două grupuri, cu o putere de 0,80 și un nivel de semnificație de 0,05. Apoi,

$$p_1 = .30;\ q_1 = 0{,}70;\ \text{alfa} = .05$$
$$p_2 = .40;\ q_2 = 0{,}60;\ \text{putere} = .80$$
$$f\ (\text{alfa, putere}) = 7{,}9 \text{ din tabel}$$
$$\frac{(.30)(.70) + (.40)(.60)}{(.40 - .30)^2} \times 7.9 = 356$$

Veți avea nevoie de 356 de persoane în fiecare grup pentru a fi sigur cu 80% că puteți detecta o diferență de la 30% la 40% la nivelul .05.

6.11 Cum se calculează dimensiunea eșantionului pentru a testa diferența dintre două mijloace

Formula pentru calcularea mărimii eșantionului pentru un test al diferenței dintre două medii, presupunând că va exista un număr egal în fiecare grup, este

n_k x-2a_ număr în fiecare grupă (MD) ^{2} ^{6p} unde o^2 este varianța erorii, *MD* este diferența minimă pe care se dorește să o detecteze și *k* depinde de nivelul de semnificație și puterea dorită. Valorile selectate ale lui *k* sunt afișate pe pagina următoare. De exemplu, pentru a detecta o

diferență a IQ-ului mediu de 5 puncte între două grupuri de persoane , unde varianța = 16 ^{2} = 25 6, la un nivel de semnificație de .05 și cu puterea de .80, am avea nevoie de

$$= \frac{7.849 \times 2(256)}{(5)^2} = 161 \text{ de persoane}$$

Nivel de semnificație	Putere	*k*
.05	.99	18.372
	.95	12.995
	.90	10.507
	.80	7.849
.01	.99	24.031
	.95	17.814
	.90	14.879
	.80	11.679

în fiecare grup sau o dimensiune totală a eșantionului de 322. Aceasta înseamnă că avem 80% șanse să detectăm o diferență mai mare sau mai mare de 5 puncte. Pentru o diferență de 10 puncte, am avea nevoie de 54 de persoane în fiecare grup.

Un set comun de parametri pentru astfel de calcule ale dimensiunii eșantionului sunt a = .05 și puterea = .80. Cu toate acestea, atunci când există mai multe comparații , trebuie să setăm o la niveluri inferioare așa cum este descris în Secțiunea 3.24 despre procedura Bonferroni. Atunci dimensiunea eșantionului nostru ar trebui să fie mai mare.

Dacă sperăm să arătăm că două tratamente sunt echivalente, trebuie să setăm diferența minimă pe care dorim să o detectăm să fie foarte mică și puterea să fie foarte, foarte mare, rezultând dimensiuni foarte mari ale eșantionului.

Pentru a calcula valorile lui *k* care nu sunt tabulate aici, cititorul este trimis la cartea *Methods in Observational Epidemiology* de Kelsey, Thompson și Evans pentru o explicație excelentă. Există programe de calculator disponibile pentru a calcula puterea pentru multe situații diferite. Un excelent este: NCSS (Software statistic al Consiliului Național pentru Studii Sociale) care poate fi obținut accesând site-ul web: www.ncss.com .

Capitolul 7
CELE MAI MULT DESPRE CALITATEA VIEI

Viața care este neexaminată nu merită trăită.

Platon, Dialoguri
428-348 î.Hr

Îmi place viața lungă mai mult decât smochinele.

Shakespeare (Anthony și Cleopatra)

Cele două citate de mai sus ilustrează cât de diferit văd oamenii calitatea vieții lor și cât de dificil este să stabilești acest concept.

O dezvoltare binevenită în cercetarea în domeniul sănătății este atenția tot mai mare acordată problemelor de calitate a vieții în studiile epidemiologice și atunci când se evaluează terapiile concurente. Un aspect cheie este măsurarea efectelor simptomelor de boală, precum și a *tratamentului* acestor simptome, asupra bunăstării, care este o stare subiectivă și relativă. Prin urmare, este destul de potrivit ca măsurarea îmbunătățirii sau a deteriorării calității vieții să se bazeze pe *percepția și auto-raportarea pacientului.* O persoană care a avut angină severă și invalidantă poate percepe o stare de bine îmbunătățită ca urmare a tratamentului dacă poate merge fără durere, în timp ce un tânăr pasionat de schi poate experimenta o deteriorare marcată dacă nu poate schia. Din acest motiv, în studiile pe această problemă, individul servește adesea drept propriul control, iar măsurile utilizate sunt *scorurile de modificare* în anumite dimensiuni ale calității vieții de la înainte până la după tratament. Cu toate acestea, rămâne important să existe un grup de control adecvat pentru a compara modificările, deoarece oamenii prezintă modificări ale acestor dimensiuni în timp, care pot fi fără legătură cu tratamentul particular evaluat.

Principiile și tehnicile descrise în această carte se aplică cercetării în orice domeniu legat de sănătate. Cu toate acestea, există anumite metode analitice care sunt deosebit de adecvate pentru investigațiile privind

 stări psihologice sau emoționale. Principiul de bază este că, dacă urmează să fie cercetare științifică, aceasta trebuie să respecte standardele științifice, ceea ce înseamnă că în primul rând *variabilele de interes trebuie cuantificate.* Din fericire, aproape orice concept legat de domeniile sănătății poate fi cuantificat dacă cineva este suficient de ingenios.

7.1 Construcție la scară

Scalele utilizate pentru măsurarea dimensiunilor calității vieții reflectă gradul de suferință cu anumite simptome sau stări psihologice, precum și gradul de satisfacție și bunăstare generală. Există multe astfel de scale disponibile, care au fost bine construite și testate pe diferite populații. Uneori, totuși, anchetatorii consideră că este necesar să-și construiască propriile scale pentru a se potrivi circumstanțelor particulare.

Există trei caracteristici ale unor astfel de scale care sunt importante: *fiabilitatea, validitatea* și *capacitatea de răspuns.*

7.2 Fiabilitate

Fiabilitatea este capacitatea de a măsura ceva în același mod de două ori. Se bazează pe presupunerea că scorul unei persoane pe o scară sau un test este compus din scorul său adevărat (dar necunoscut) plus o componentă care este supusă variației din cauza erorii (prin care înțelegem variabilitatea aleatorie).

Fiabilitatea unei scale este legată de *repetabilitatea acesteia* sau de cât de apropiate sunt răspunsurile la două administrări ale scalei. Pentru a măsura cât de apropiate sunt acestea putem calcula coeficientul de corelație dintre cele două administrări ale scalei la aceiași subiecți. Dar adesea nu putem acorda aceeași scară pacienților noștri de două ori exact în aceleași circumstanțe , deoarece, în realitate, un pacient care răspunde de două ori la aceleași întrebări ar răspunde diferit, fie pentru că a intervenit ceva între cele două ocazii, fie pentru că și-a amintit răspunsurile anterioare, fie doar pentru că există o variabilitate inerentă în modul în care se simte. Următorul lucru cel mai bun ar fi să dai două scale echivalente la

același grup, dar are și problemele sale. De unde știm că cele două scale sunt cu adevărat echivalente?

Din fericire, există diverse măsuri ale ceea ce numim „coerență internă" care ne oferă fiabilitatea unei scale sau a unui test. Cel mai comun se numește alfa lui Cronbach. Există multe pachete software pentru computere personale care calculează cu ușurință al pha lui Cronbach, inclusiv SPSS, SAS, STATA și multe altele. Astfel, nu este necesar să îl calculați singur, dar următoarea explicație indică ce înseamnă cu adevărat și cum să o interpretați.

$$k\text{-}1 \qquad \frac{\text{varianța scalei totale - suma variațiilor itemilor individuali}}{\text{variația scării totale}}$$

Varianta este abaterea standard la pătrat. Secțiunea 3.4 arată cum se calculează. (Când vorbim despre varianță aici, ne referim de fapt la varianța populației, dar ceea ce folosim cu adevărat sunt estimări ale varianței populației pe care le obținem din eșantionul special de persoane pe care dezvoltăm testul sau scara, deoarece evident că nu putem măsura întreaga populație.)

Această formulă este într-adevăr o măsură a cât de omogene sunt elementele de scară, adică în ce măsură măsoară același lucru. Dacă aveți o scară care este compusă din mai multe subscale diferite, fiecare măsurând lucruri diferite, atunci alfa lui Cronbach ar trebui utilizată pentru fiecare dintre subscale separat, mai degrabă decât întreaga scară. Alfa lui Cronbach oferă limita inferioară pentru fiabilitate. Dacă este mare pentru întreaga scală, atunci știți că scala este de încredere (repetabilă, foarte corelată cu scorurile „adevărate", dar necunoscute). Dacă obțineți un alfa scăzut pentru întreaga scară, atunci fie nu este de încredere, fie măsoară mai multe lucruri diferite.

Fiabilitatea poate fi privită și ca o măsură a corelației și, de fapt, reflectă corelația medie între itemii unei scale, luând în considerare numărul total de itemi. O altă modalitate de a obține fiabilitate este din formula Spearman-Brown, care este

$$k(\text{corelația medie între toți itemii}) \quad k \; (\text{media } r)$$

$$1 + (k-1)\,\text{corelația medie între toate elementele} \quad 1 + (k-1)(r_{\text{media}})$$

După cum indică această formulă, un test sau o scară mai lungă este, în general, mai fiabil dacă elementele suplimentare măsoară același lucru. Pe de altă parte, scalele mai scurte sunt mai acceptabile pentru pacienți. Un alfa peste .80 este considerat foarte bun, iar uneori subscalele sunt acceptabile cu al pha peste .50, mai ales atunci când există un număr mare de subiecți (peste 300), dar ar trebui luat în considerare în contextul celorlalte calități psihometrice ale scalei.

Există și alte măsuri de fiabilitate. Psihometria este un domeniu specializat și complex și există multe cărți excelente pe acest subiect, de exemplu, *Scale de măsurare a sănătății* de Streiner și Norman.

7.3 Valabilitate

Valabilitatea se referă la gradul în care testul măsoară ceea ce ar trebui să măsoare. O situație ideală ar fi aceea în care există un criteriu extern în funcție de care să judeci instrumentul de măsurare, un „etalon de aur". De exemplu, dacă s-ar putea demonstra că anxietatea măsurată pe o scară se corelează mai bine cu un rezultat definibil și convenit în mod obiectiv decât anxietatea măsurată pe o a doua scară, s-ar putea spune că prima scală este mai validă. (Acest lucru se numește „validitate de criteriu").

Din păcate, în problemele de calitate a vieții, în general, nu există criterii externe. O persoană poate simți că este nefericită, dar funcționează la un nivel înalt. Însăși ideea de calitate a vieții este conceptual subiectivă. Oricum a cui este calitatea vieții?

Prin urmare, de multe ori trebuie să ne bazăm pe validitatea conținutului, care este un amestec de bun simț și proprietăți psihometrice tehnice. Dacă vrem să știm dacă cineva se simte deprimat, am putea întreba: „Te simți foarte trist?" mai degrabă decât: „Te simți atletic?" Cu toate acestea, chiar și asta nu este atât de simplu, deoarece ceea ce cineva care nu este un expert în depresie

ar putea considera în mod evident irelevant, cum ar fi tulburările de somn, este unul dintre cele mai puternice semne de depresie.

Desigur, dacă există un criteriu extern în raport cu care să se valideze o scală, acesta ar trebui utilizat. Dar chiar și validitatea conținutului poate deveni mai obiectivă, de exemplu prin formarea unui grup de experți care să emită judecăți asupra validității conținutului articolelor. Pentru a testa acordul dintre judecători, poate fi utilizat coeficientul kappa, așa cum este descris în Secțiunea 3.3.

7.4 Receptivitatea

Capacitatea de răspuns a unei scale este o măsură a cât de bine poate detecta schimbările ca răspuns la o intervenție. Reactivitatea sau sensibilitatea unei scale poate fi evaluată în mai multe moduri diferite și nu există un consens cu privire la care este cel mai bun. Unele concepte înrudite sunt descrise mai jos, care se referă la situația în care vă uitați la schimbarea măsurilor pre- și post-tratament.

(1) Utilizarea *scorurilor de modificare* (pre-post) este adecvată atunci când variabilitatea dintre pacienți este mai mare decât variabilitatea în interiorul pacienților . În general, scorurile de modificare pot fi utilizate în siguranță dacă

$$\frac{n_{\text{între pacienți}}}{n - n_{\text{între}}\, n_{\text{eroare}}^{1}} > 0.5$$

Eroarea n^{2} între pacienți și n^{2} poate fi obținută dintr-o analiză a variației scorurilor unui grup de pacienți care au replicat măsuri, astfel încât să puteți estima varianța datorată erorii.

(2) Un *coeficient de sensibilitate* la schimbare datorată unui tratament este

$$\frac{n_{\text{schimba}}^{\_\,2}}{n + n^{2}}$$

$n_{schimbare}$ n_{eroare}

Pentru a obține $_{eroarea}o^2$, trebuie să faceți o analiză a varianței măsurilor repetate pe aceleași subiecte. Sunt disponibile programe de calculator. Explicații detaliate ale acestui lucru apar în texte mai avansate.

(3) *Mărimea efectului* este pur și simplu modificarea scalei de la înainte la după tratament, împărțită la abaterea standard la momentul inițial. Abaterea standard este un indice al variabilității generale a scorurilor în rândul grupului de persoane din studiu. Se poate măsura magnitudinea modificării medii a scorurilor după un anumit tratament, determinând ce procent din „variația de fundal" reprezintă această modificare. *Dimensiunea efectului =*

$$\frac{\textit{scorul mediu al schimbării}}{\textit{abaterea standard a scorurilor de bază (sau pretratament).}}$$

(4) O *măsură a receptivității* propusă de Guyatt și colab. [25] este

$$\frac{\textit{scorul mediu al schimbării}}{\textit{abaterea standard a scorurilor de schimbare pentru „subiecții stabili"}}$$

„Subii stabili" sunt greu de definit, dar ceea ce sugerează acest lucru este că poate fi utilizat un grup de control care nu primește intervenția sau primește placebo. Apoi se poate folosi abaterea standard a scorurilor de schimbare din grupul de control ca numitor în termenul de mai sus.

Variabilitatea scorurilor de schimbare în grupul de control (sau într-un grup de subiecți stabili) poate fi privită ca „variabilitatea de fundal " a *schimbărilor* , iar instrumentul de măsurare răspunde la gradul în care poate detecta schimbări dincolo de această variabilitate de fond.

(5) Când se evaluează schimbarea datorată tratamentului, ar trebui să existe *întotdeauna* un grup de control (adică, un grup fără intervenție sau placebo) pentru comparație , deoarece schimbarea poate apărea și la pacienții de control, iar întrebarea de interes este dacă schimbarea

înaintea tratamentului în grupul de tratament depășește schimbarea „de fond” în grupul de control. Dacă utilizați dimensiunea efectului ca măsură, atunci ar trebui să comparați dimensiunea efectului în grupul de tratament cu dimensiunea efectului în grupul de control.

Un exemplu numeric al acestor concepte este oferit în Anexa G.

7.5 Câteva capcane potențiale

a. Multiplicitatea variabilelor

Cercetarea calității vieții se ocupă adesea de o mare cantitate de variabile. Să presupunem că un investigator examinează efectele unui medicament pentru tratarea hipertensiunii și îl compară cu placebo. Investigatorul poate avea câteva sute de elemente pentru a evalua diferite simptome fizice și psihologice și efecte secundare. Dacă s-ar compara cele două grupuri prin testul t pe fiecare dintre itemi, la nivelul de semnificație *p = 0,05, ne-am* aștepta ca aproximativ 5% dintre aceste teste să producă un rezultat semnificativ doar din întâmplare. Probabilitatea exactă este dificil de determinat, deoarece unele dintre aceste comparații ar fi corelate în virtutea faptului că aceiași pacienți răspund la toate, adică răspunsurile nu sunt independente. Dar, în orice caz, dacă anchetatorii aleg doar elementele semnificative și concluzionează că există efecte ale medicamentului, este posibil să comită erori de tip I, adică să respingă incorect ipoteza nulă.

De aceea, este important să folosiți scale care măsoară anumite constructe sau să grupați itemii într-un mod semnificativ din punct de vedere clinic. De exemplu, s-ar putea dori să măsoare depresia, anxietatea, ostilitatea, bunăstarea (fiecare constă din mai multe elemente). Pe de altă parte, anumite medicamente pot fi legate de simptome foarte specifice, cum ar fi coșmarurile, iar acest lucru ar putea trebui evaluat printr-un singur element care să întrebe despre frecvența coșmarurilor.

Ideea este că cercetarea privind calitatea vieții ar trebui, în general, să fie condusă de unele ipoteze specifice. În caz contrar, devine o „ expediție de pescuit” care doar pescuiește în jur pentru orice poate găsi semnificativ. Trebuie remarcat faptul că „expedițiile de pescuit” pot fi utile pentru a genera ipoteze care apoi trebuie testate într-un alt studiu.

b. Generalizare

O altă problemă importantă este extrapolarea rezultatelor către alte populații decât cea din care a fost extras eșantionul de studiu. Efectele calității vieții pot fi diferite la bărbați decât la femei, la persoanele mai tinere decât la persoanele în vârstă și pot diferi în funcție de grupurile etnice și culturale. Ar trebui să fii atent în a face generalizări.

c. Necesitatea unor standarde riguroase de cercetare

Unii oameni consideră că măsurile de calitate a vieții sunt „soft". Ceea ce înseamnă în general este că ei cred că astfel de măsuri sunt subiective, variabile și poate lipsite de sens. Este un nonsens și, în măsura în care este adevărat în unele studii, reflectă insuficiența cercetătorului, nu a subiectului. Aceste măsuri *ar trebui să fie subiective* din perspectiva pacientului, deoarece reflectă percepția subiectivă a pacientului despre bunăstare sau suferință. *Cercetătorul este cel care nu ar trebui să fie subiectiv* și care nu trebuie să fie dacă urmează principiile cercetării. Variabilitatea măsurătorilor calității vieții nu este mai mare decât în multe măsuri fiziologice și, în orice caz, face parte din esența unor constructe ale calității vieții. În ceea ce privește sensul, aceasta este o problemă filozofică, nu științifică . Din punct de vedere științific, „sensul" ar trebui definit - operațional. Cercetarea privind calitatea vieții ar trebui să adere la principiile oricărei cercetări bune, iar abordarea generală este aceeași ca pentru orice investigație științifică:

(1) formularea unei ipoteze testabile;
(2) cuantificați variabila (sau variabilele) dependentă;
(3) selectați un design de studiu care poate răspunde la întrebarea pe care ați pus-o;
(4) cuantificarea variabilelor independente;
(5) controlul potențialilor factori de confuzie (prin proiectarea studiului și/sau analiza datelor);

(6) planificați o dimensiune a eșantionului care vă va oferi suficientă

putere pentru a detecta dimensiunea efectului de interes;
(7) încercați să vă asigurați că minimizați părtinirea sistematică;
(8) colectarea datelor, acordând o mare atenție controlului calității;
(9) analiza datelor folosind tehnici statistice adecvate; şi
(10) face inferențe în concordanță cu punctele forte și limitările Studiul.

Capitolul 8
MAI MULT DESPRE EPIDEMIOLOGIA GENETICĂ

Să presupunem atunci că mintea este, după cum spunem, hârtie albă (tabula rasa), lipsită de toate caracterele fără nicio idee; cum se face sa fie mobilat? De unde provine acel vast magazin, pe care fantezia ocupată și nemărginită a omului l-a pictat cu o varietate aproape nesfârșită?... La aceasta răspund, într-un cuvânt, Din experiență: în aceea că toată cunoștințele noastre se întemeiază.....

John Locke Un eseu despre înțelegerea umană (1689)

8.1 O nouă eră științifică

Suntem departe de a crede că mintea este o „tabula rasa", o tablă goală. Știm acum că multe sunt de fapt înnăscute, adică sub influență genetică. Scopul acestui capitol este de a-i ajuta pe cei care doresc să citească literatura în expansiune rapidă în epidemiologia genetică. Astfel, este o privire de ansamblu asupra proiectelor de bază și a statisticilor utilizate în acest domeniu; nu este cuprinzător și nici nu este foarte tehnic. Anexa H oferă descrieri de bază și definiții ale procesului intracelular sub controlul genelor, pentru cei care nu sunt familiarizați cu zona.

Accentul cercetării epidemiologice a evoluat pe măsură ce s-au înregistrat progrese paralele în alte domenii ale medicinei și științei fundamentale. În epoca în care bolile infecțioase erau răspândite, epidemiologia s-a preocupat de identificarea surselor de infecție și a metodelor de transmitere, în mare parte prin munca de teren. Pe măsură ce agenții infecțioși au fost descoperiți, odată cu îmbunătățirea stării sanitare și a stării de sănătate, bolile cronice, cum ar fi bolile de inimă și cancerul, au devenit principalele cauze de deces și dizabilități în lumea dezvoltată și au ajuns să fie obiectivele principale ale cercetării epidemiologice. (Acum, că noi boli infecțioase apar din nou, această parte a epidemiologiei câștigă din nou proeminență.)

Obiectivul epidemiologiei bolilor cronice a fost identificarea factorilor de risc pentru aceste boli. Această parte a poveștii a avut un mare succes de

sănătate publică. Știm acum, datorită studiilor epidemiologice, care sunt factorii de risc majori modificabili pentru bolile cardiovasculare: - hipertensiune arterială, colesterol ridicat și HDL scăzut, fumat, supraponderalitate și inactivitate. Provocarea noastră acum este să găsim modalități de a face schimbări ale stilului de viață în populație, care vor scădea și mai mult ratele bolilor cardio-vasculare. De asemenea, cunoaștem multe dintre expunerile legate de cancer, dar nu la fel de cuprinzător ca pentru bolile de inimă.

În acest moment istoric din punct de vedere științific, pe măsură ce știința intră pe deplin în era *genomică și proteomică* , epidemiologia intră într-o nouă fază a activității de cercetare: *epidemiologia moleculară* . Aceasta este căutarea de biomarkeri din sânge sau țesut și polimorfisme genetice (variante) care sunt asociate sau predispun la boli. De ce este acesta diferit de orice alt factor de risc investigat în epidemiologie? În multe privințe, nu este, mai ales în ceea ce privește biomarkerii din sânge, dar în epidemiologia genetică există modele de studiu și metode de analiză statistică care sunt destul de diferite. Un aspect cu adevărat nou al - epidemiologiei moleculare și genetice este adevărata colaborare a oamenilor de știință , clinicienilor și epidemiologilor. De prea mult timp disciplinele au urmat căi de cercetare separate, iar oamenii de știință citesc mai ales reviste științifice din domeniul lor. Dar epidemiologia moleculară nu poate continua cu succes fără interfața oamenilor de știință de laborator și a cercetătorilor populației.

8.2 Prezentare generală a epidemiologiei genetice

Epidemiologia genetică urmărește să identifice genele legate de boală și să evalueze impactul factorilor genetici asupra sănătății și bolii populației. Iată o privire de ansamblu asupra strategiei adesea folosite pentru a studia determinanții genetici ai bolii. Mai întâi putem dori să stabilim dacă boala *se transmite în familii* . Dacă nu este familială, nu este probabil să fie ereditară; dacă este familial , poate fi sau nu din cauza unor factori genetici (medii se desfășoară și în familii). Apoi, vrem să vedem dacă genele contribuie la transmiterea familială. O metodă pentru a determina acest lucru este studierea gemenilor (descris în Secțiunea 8.3). Dacă stabilim că *boala este ereditară,* am dori să *localizăm și să identificăm* genele

implicate.

Ca un prim pas, putem dori să aflăm unde sunt localizate genele care contribuie la tulburare. Acest lucru poate fi realizat prin studii de legare a scanării genomului asupra indivizilor afectați de boală și a familiilor acestora (descris în Secțiunea 8.4). *Studiile de legare* pot identifica regiuni de pe cromozom care sunt susceptibile de a adăposti genele bolii. Odată ce am identificat una sau mai multe astfel de regiuni, putem căuta să vedem ce gene se știe că locuiesc în acele regiuni. Putem apoi testa aceste gene folosind *studii de asociere la indivizi neînrudiți* (descrise în Secțiunea 8.6) pentru a determina dacă vreo variantă (numită și *alele*) ale acestor gene este asociată cu boala.

Deci, există o varietate de modele și teste statistice care pot fi utilizate pentru a defini baza genetică a unei boli, inclusiv: 1) *studii pe gemeni* pentru a determina dacă boala are o componentă ereditară; 2) *studii de legătură* pentru identificarea și localizarea regiunilor cromozomilor care conțin gene implicate în boală; și 3) *studii de asociere* pentru a determina dacă anumite variante genetice sunt asociate cu boala, pentru a examina modul în care acestea interacționează cu mediul și pentru a determina modul în care afectează sănătatea populației. Vom limita discutia la cateva modele destul de simple care vor da savoarea subiectului. Cititorii interesați de mai multă profunzime sunt referiți la multe mai multe scrieri tehnice pe acest subiect, dar în special la excelentul Raport special despre genetică de Ellsworth și Manolio. [28,30,31]

8.3 Studii pe gemeni

Pentru a stabili o influență genetică asupra bolii, putem căuta să vedem dacă aceasta rulează în familii. Dar ceva care este familial nu este neapărat ereditar. De exemplu: părinții obezi au copii obezi din cauza geneticii sau din cauza nutriției și a nivelurilor de activitate care se transmit de la părinți la copii? Ceea ce vrem să știm este dacă și în ce măsură fenotipul (ceea ce observăm la persoană, de exemplu, obezitatea) este afectat de factori genetici.

O modalitate de a evalua influența eredității este din studiile pe gemeni. Gemenii identici (monozigoți – care provin din același ovul fertilizat)

împărtășesc 100% din genele lor, în timp ce gemenii fraterni (dizigoți – care provin din două ouă fertilizate) împărtășesc doar 50%, la fel cum fac frații negemeni. O modalitate de a estima puterea influențelor genetice este prin intermediul unui *indice de ereditare* h^2. Aceasta este definită în mod obișnuit ca de două ori diferența dintre corelația pentru acea trăsătură între gemenii monozigoți minus relația cor la gemenii dizigoți sau:

$$h^2 = 2(r_{raz} - rj$$

Luați în considerare tensiunea arterială. Dacă variația stării sau trăsăturii investigate ar fi complet atribuită variației genetice, atunci fiecare membru al unei perechi de gemeni monozigoți ar fi afectat în mod egal (fiecare membru ar avea aceeași tensiune arterială) și corelația dintre gemenii monozigoți ar fi 1,0; corelația la gemenii dizigoți ar fi totuși .50.

În acest caz, h^2 ar fi 2(1- 0,5) = 1,0 sau 100%. Dacă condiția nu este complet ereditară în populație, atunci $r_{mz} = r_{dz}$ și $h^2 = 0$. Deoarece bolile și trăsăturile sunt în general parțial ereditare, h^2 se află undeva între 0 și 1.0.

Dacă vorbim de variabile continue, ne putem gândi la ereditate în termeni de coeficienți de corelație. Dacă vorbim de variabile categorice, putem vorbi de rate de concordanță, unde

$$h^2 = \frac{\%\ \text{gemeni monozigoți concordanți pentru boală} - \%\ \text{dintre gemenii dizigoți concordanți}}{1 - \%\text{din gemeni dizigoti concordanti}}$$

Unele estimări aproximative raportate [32,33,34,35] ale ereditabilității sunt: 1,0 pentru boala Huntington (aceasta deoarece boala Huntington este o tulburare cu o singură genă, moștenită într-un mod dominant de transmitere și complet penetrant - acest termen este definit la sfârșitul acestui capitol); .60 pentru alcoolism; .40 pentru trăsătura de personalitate a conștiinciozității; .35 pentru cancerul colorectal;.26 pentru scleroza multipla; și .80 pentru schizofrenie.

Este important să ne amintim că ereditatea nu măsoară cât de mult din boala unui individ este atribuită geneticii; mai degrabă ne spune ce proporție din variabilitatea populației în fenotip este rezultatul variației genelor din populație. Deci este o măsură aplicabilă unei populații, nu unui individ. Dacă aveți oameni care trăiesc exact în același mediu, atunci orice variație pe care o întâlniți în fenotip trebuie să se datoreze factorilor genetici,

deoarece nu există nicio variație de mediu. Într-un astfel de caz, dacă toți factorii de mediu sunt constanți pentru populație, ereditabilitatea ar fi de 100%. Deci, există unele limitări ale acestei măsuri, dar ne oferă o idee în ce măsură variația genetică contribuie la variația fenotipică într-o populație . Cu toate acestea, ereditabilitatea nu ne spune nimic despre genele responsabile pentru acea variație, pe ce cromozomi se află, unde se află pe cromozom sau ce polimorfisme sunt implicate.

8.4 Studii de legături și asocieri

Dacă știm că o boală este ereditară, ne putem întoarce acum la sarcina de a identifica efectiv genele implicate. Cele mai multe tulburări care sunt studiate de epidemiologi (de exemplu, boli cardiovasculare, tulburări psihice , forme comune de cancer) sunt considerate tulburări „complexe". Adică, spre deosebire de tulburările cu o singură genă sau mendeliană, cum ar fi fibroza chistică sau boala Huntington, se crede că aceste boli rezultă din contribuția mai multor sau mai multor gene care interacționează cu - factorii de risc de mediu. Acest lucru poate face identificarea efectului unei gene individuale o sarcină destul de dificilă. Efectul unei anumite alele din acea genă poate fi destul de mic. Este un pic ca și cum ai căuta proverbialul ac în carul de fân. Cu toate acestea, genele care contribuie la boli sunt descoperite și există anumite strategii care sunt folosite în căutare.

Unde în genom căutăm genele care conferă susceptibilitate la boală? O modalitate de a răspunde la această întrebare este utilizarea *analizei legăturilor genetice* .

a) Analiza legăturii se bazează pe fenomenele de încrucișare și recombinare care apar în timpul procesului de meioză când se formează celulele sexuale (sperma și ovulul). Fiecare persoană are două copii ale fiecăruia dintre cei 23 de cromozomi care alcătuiesc genomul : o copie este moștenită de la mamă și una de la tată. În timpul formării spermatozoidului și a ovulelor, aceste 23 de perechi de cromozomi se aliniază și schimbă segmente de material genetic într-un proces cunoscut sub numele de *crossing over*. Această recombinare are loc în unul sau mai multe locuri de-a lungul cromozomului. *Cu cât doi loci sunt mai apropiați pe un cromozom, cu atât este mai puțin probabil să apară un eveniment de re-*

combinație între ei și, prin urmare, cu atât este mai probabil ca aceștia să fie moșteniți împreună. Loci care tind să fie comoșteniți se spune că sunt legați genetic. Putem folosi acest fapt pentru a estima distanța dintre doi loci sau markeri genetici (un marker genetic este o bucată de ADN a cărei locație cromozomială este cunoscută). Distanța fizică dintre doi markeri este în sens invers legată de cât de des sunt moșteniți în comun de-a lungul generațiilor într-o familie.

b) Distanța *dintre doi loci este uneori măsurată în centi Morgans* . Un centiMorgan (cM) este o unitate de distanță de-a lungul unui cromozom, dar nu în sensul obișnuit al distanței fizice . Este într-adevăr o măsură de probabilitate care este o reflectare a distanței fizice; reflectă probabilitatea ca doi markeri sau loci să fie separați (sau segregați) prin încrucișare în timpul meiozei. Dacă cei doi markeri sunt foarte apropiați unul de celălalt, nu se vor separa (spunem că sunt „legați"); dacă sunt departe unul de celălalt, este mai probabil să se încrucișeze și *materialul genetic se recombină în timpul meiozei* . Apoi, acest ADN recombinat este transmis - descendenților. Doi loci sunt la un centimorgan unul de celălalt dacă probabilitatea ca ei să fie separați prin încrucișare este de numai 1% (o dată la o sută de meioze). S-a estimat că există aproximativ 1 milion de perechi de baze într-un interval de 1cM. Locii care sunt departe unul de celălalt, să zicem 50 cM, vor fi moșteniți independent unul de celălalt, așa cum ar fi dacă ar fi pe cromozomi diferiți. *Scopul studiilor de legătură este de a localiza gena de susceptibilitate a bolii pentru a se afla într-o anumită regiune a cromo- unii* .

c) Așadar, am putea începe căutarea unei gene a bolii prin colectarea familiilor afectate de boală și efectuând o analiză a legăturii folosind markeri distanțați la intervale (să zicem la 10 cM unul de celălalt) de-a lungul întregului genom. Aceasta este cunoscută ca *o scanare a întregului genom.* Dacă găsim un marker care pare a fi co-transmis cu boala, am avea dovezi că markerul este legat genetic de o genă a bolii. Cu alte cuvinte, este probabil să existe o genă pentru boală în aceeași regiune cu markerul legat (există un ponei

acolo undeva. §)

d) După ce am găsit o regiune cromozomială legată de boală, am putea încerca să restrângem regiunea prin genotiparea și testarea markerilor suplimentari în acea regiune (să zicem la intervale de 1 cM). Cu toate acestea, chiar și această regiune relativ mică poate conține multe gene.
e) Următorul nostru pas ar putea fi să analizăm genele despre care se știe că locuiesc în această regiune. Ne-ar interesa în special genele care au o legătură plauzibilă cu boala de interes (acestea ar fi bune *„gene candidate").* De exemplu, dacă studiem diabetul, genele care produc proteine implicate în metabolismul glucozei ar fi „gene candidate" importante.
f) Acum putem vedea dacă anumite alele (variante) ale genelor din acea regiune cromozomială sunt asociate cu boala. Acest lucru se poate face prin:

   1. *Studii de asociere care utilizează metode caz-control la - persoane neînrudite,* examinând dacă o alelă este mai frecventă în cazuri decât la martori (descris în secțiunea 8.6);
   2. *Studii de asociere în familii* pentru a vedea dacă o alelă este transmisă mai frecvent la cazuri decât se aștepta întâmplător (descris în Secțiunea 8.7).

Deci, analiza legăturii ne spune că o anumită locație a markerului este în apropierea unei gene de susceptibilitate a bolii; analiza de asociere ne spune că o anumită alelă a unei gene sau a unui marker este mai frecvent

---

§ Aceasta se referă la povestea unei perechi de gemeni. Unul era optimistul și unul pesimist. Psihologul a încercat să-i vindece de distorsiunile respective și așa l-a băgat pe geamănul pesimist într-o cameră plină de jucării. Copilul a plâns - consolator, explicându-i că plânge pentru că nu va putea niciodată să aleagă o jucărie cu care să se joace din toată această grămadă mare. Psihologul l-a pus pe geamănul optimist într-o cameră plină cu gunoi de cal, iar când s-a întors o jumătate de oră mai târziu, copilul râdea și bătea din palme de bucurie. „Cum de ești atât de fericit aici", a întrebat psihologul, iar geamănul optimist a răspuns: „Îmi dau seama că, cu tot acest gunoi de grajd de cai, trebuie să fie un ponei acolo pe undeva."

Datele de moștenire observate cu probabilitate, având în vedere nicio legătură

moștenită de persoanele cu boală.

8.5 Scor LOD: Statistică de legătură

Statistica clasică folosită pentru a evalua puterea dovezilor în favoarea - legăturii unui marker genetic și a genei de susceptibilitate a bolii este scorul LOD (log $_{10}$ al șanselor în favoarea legăturii). Acesta va fi descris doar în principiu, pentru a ajuta la interpretarea articolelor epidemiologice care se ocupă de genetică. Calculele reale sunt complexe și necesită pachete speciale de programe.

Principiul care stă la baza scorului LOD este descris în secțiunea anterioară: dacă avem doi loci - de exemplu, un marker și o genă a bolii - cu cât sunt mai aproape de un cromozom, cu atât este mai mică probabilitatea ca ei să fie separați de un eveniment de recombinare în timpul meiozei și cu atât este mai probabil ca ei să fie co-moșteniți de descendenți.

Probabilitatea recombinării, numită fracțiune de recombinare , este notă cu simbolul 9 și depinde de distanța dintre genă și marker. *Dacă nu există recombinare și gena și markerul sunt complet legate, atunci fracția de recombinare este 0.* Valoarea maximă de 9 este .5 (dacă gena și markerul au fost moștenite independent, atunci probabilitatea ca markerul să fi fost transmis, dar nu gena, este egală cu probabilitatea ca gena să fi fost transmisă, dar nu markerul, este egală cu .50).

Deci, dacă doriți să știți dacă există o legătură, trebuie să estimăm cât de probabil este ca 9 să fie mai mic de .5, având în vedere datele pe care le-am observat. Pentru aceasta, folosim raportul de probabilitate, care, după cum vă amintiți din capitolul 2, este raportul dintre probabilitatea simptomelor observate, dată boală împărțită la probabilitatea simptomelor observate fără boală. În acest caz:

$$LR = \frac{\textit{Probabilitatea datelor de moștenire observate, dată de legătură}}{}$$

Ipoteza nulă aici nu este nicio legătură (sau fracțiunea de recombinare 9 = .5) și ipoteza alternativă este legătura (sau 9 <.5). Dacă respingem nulul, „acceptăm" ipoteza alternativă. Statisticile de testare folosite pentru a vedea dacă avem suficiente date pentru a concluziona legătura este *scorul LOD,*

care este log 10 (LR.). Pentru tulburările mendeliane (o singură genă), un scor lod de 3 a fost în mod tradițional pragul pentru declararea unei legături semnificative, deși pentru tulburările complexe au fost recomandate praguri mai mari (3,3 - 3,6). Un scor LOD de 3,0 indică 103 cote în favoarea legăturii în comparație cu nicio legătură, adică 1000:1 cote în favoarea legăturii.

Din motive complexe dincolo de scopul acestei cărți (dar descrise în referințele de la sfârșit), un scor LOD poate fi tradus în probabilitate prin înmulțirea lui cu constanta 4,6: LOD x 4,6 este distribuit ca chi-pătrat cu 1 grad de libertate. (4,6 este de 2 ori logaritmul natural de 10.) Astfel, un LOD de 3,0 este echivalent cu un chi-pătrat de 3 x 4,6 =13,82 și corespunde cu *p* = .0002. Datele de moștenire pentru analizele de legături pot proveni din studii de genealogie a familiei, de la frați sau alte grupuri de familie.

Analiza de legare a scorului LOD este uneori denumită analiză de legare „parametrică" deoarece necesită specificarea anumitor parametri (de exemplu, frecvența alelelor de boală și marker, modul recesiv vs. dominant de moștenire, penetranța genei bolii). Atunci când acești parametri sunt cunoscuți sau pot fi aproximați, analiza parametrică a scorului LOD este cea mai puternică metodă de analiză a legăturilor. Acest lucru poate fi adevărat pentru boala mendeliană (o singură genă), dar pentru multe tulburări complexe, acești parametri nu sunt cunoscuți. Metodele de legare „nonparametrice" (cunoscute sub numele de abordarea partajării alelelor) sunt adesea folosite pentru a studia tulburările complexe, deoarece nu necesită cunoașterea modului de moștenire sau a altor parametri genetici. Există o serie de statistici disponibile, descrise în textele mai avansate.

8.6 Studii de Asociere

În comparație cu analiza legăturii, studiile de asociere sunt mai strâns asemănătoare cu studiile epidemiologice tradiționale. Ele pot fi utilizate pentru a evalua o genă candidată pe care o investigăm atunci când datele anterioare, în general din studiile de legătură, au sugerat că o anumită genă este implicată și poate chiar o anumită variantă a genei. În aceste studii, cercetătorii sunt interesați să afle dacă există vreo asociere între o anumită alelă a unei gene polimorfe și boala în cauză. (Un polimorfism este o variație la un anumit loc de pe cromozom). În scopul acestei discuții, vom

presupune că polimorfismele pe care le analizăm sunt *SNP-uri (- polimorfisme cu un singur nucleotide)* sau variante într-o singură dintre bazele ATCG la un anumit locus (vezi Anexa H pentru detalii despre SNP).

Deci, să spunem că, la un anumit SNP, unii oameni au alela A și alți oameni au alela G. Vrem să știm dacă persoanele cu boala au mai multe șanse să spună, alela A decât alela G. Putem face studii de asociere caz-control, luând cazuri care sunt afectate de boală și controale neînrudite care nu sunt. Testăm apoi statistic dacă proporția de indivizi afectați care au alela A este mai mare decât proporția de control care au alela A.

Dacă unitatea de observație este persoana, putem analiza datele după genotip (adică comparați proporțiile de cazuri și martori cu o alele de la mamă și una de la tată, rezultând genotipul lor de AA, AG, vs GG,). Putem folosi teste statistice obișnuite ale diferențelor dintre proporții sau regresii logistice multiple (a se vedea secțiunea 4.16) pentru a determina raportul de șanse conectat cu alela în cauză și putem testa interacțiunile genă-mediu prin includerea unui termen de interacțiune a prezenței alelei și a unui factor de mediu, cum ar fi fumatul. Studiile de asociere pot fi mai puternice decât analiza legăturii pentru detectarea genelor cu efect modest, făcându-le o abordare atractivă pentru studiul tulburărilor complexe care se așteaptă să implice mai multe gene cu efect individual relativ mic.

O problemă potențială legată de astfel de studii este alegerea grupurilor de control. Diferențele etnice în prevalența alelei, care pot fi fără legătură cu prezența bolii și care au apărut deoarece alela a fost transmisă de-a lungul generațiilor în cadrul grupurilor etnice, pot confunda rezultatele noastre. Acest fenomen este denumit *„ stratificarea populației”.* Prin urmare, în primul rând, ar trebui să știm
despre distribuția de fond a acestor alele în diferite subgrupuri etnice de populație.

De exemplu, să presupunem că varianta A este mai frecventă la caucazieni și varianta G este mai frecventă la afro-americani. (Asemenea diferențe între grupuri se pot întâmpla din cauza diferiților progenitori din diferitele grupuri.) Dacă se întâmplă ca grupul nostru de boală să aibă mai mulți caucazieni și grupul nostru de control fără boală să aibă mai mulți afro-americani, atunci am putea descoperi că alela A este mai frecventă printre cei cu boală decât la cei fără boală, dar s-ar putea să fie doar o reflectare a faptului că am avut mai mulți caucazieni și purtători de control A sunt mai probabil ca purtători caucazieni decât A. lele. Deci am avea o

constatare fals pozitivă din cauza compoziției etnice a celor două grupuri. Dar etnia nu este atât de simplu determinată, deoarece atât caucazienii, cât și afro-americanii au origini etnice multiple și pot avea progenitori diferiți și modele diferite de alele care apar în mod normal.

Astfel, dacă găsim o asociere între o anumită alelă și boala pe care o studiem, aceasta poate fi din unul dintre cele patru motive: (1) ar putea fi un fals pozitiv din cauza întâmplării (eroare de tip I); (2) ar putea fi un fals pozitiv din cauza confuziei din cauza stratificării populației; 3) s-ar putea ca alela să fie în „dezechilibru de legătură" cu alela adevărată a bolii, ceea ce înseamnă că alela pe care am găsit-o mai frecvent în cazuri decât la martori, este situată fizic suficient de aproape de alela adevărată a bolii, încât cele două alele tind să fie moștenite împreună și să apară concomitent la indivizii afectați; 4) există într-adevăr o adevărată asociere cauzală a alelei pe care am studiat-o și a bolii. (Așa cum am spus mai înainte, genetica și viața nu sunt simple.)

8.7 Teste de dezechilibru al transmisiei (TDT)

Un test statistic al legăturii și al asocierii este *testul de dezechilibru al transmisiei* . Acesta este un test de asociere bazat pe familie care evită confuzia din cauza stratificării populației prin examinarea transmiterii - alelelor *în cadrul* familiilor. Dacă un marker și un polimorfism într-o genă de susceptibilitate la boală sunt în *dezechilibru de legătură,* înseamnă că

ele sunt atât de strâns legate încât alelele specifice ale markerului tind să fie moștenite împreună cu alele specifice ale genei. Să luăm din nou în considerare un SNP la care există două alele posibile, A și G. Un părinte care este heterozigot la acest SNP (adică are genotipul AG) poate transmite fie un A sau un G copilului. În ipoteza nulă a lipsei de legături și a asocierii, probabilitatea de a transmite oricare dintre aceste alele este de 50%. Dacă observăm transmiterea alelei A semnificativ mai des decât așteptările întâmplătoare (adică mai mult de 50% din timp) către descendenții afectați, atunci ajungem la concluzia că alela A este asociată cu boala.

Statisticile de bază sunt destul de simple. Unitatea de analiză este *transmisia unei alele* . Iată un exemplu folosind triouri. Un trio este un descendent afectat și ambii părinți. Există două transmisii posibile ale unei alele într-un anumit locus - una de la mamă și una de la tată. În diagrama de mai jos construim un tabel 2x2 și numărăm numărul de alele transmise care aparțin fiecărei celule a tabelului.

Trio 1: deoarece copilul afectat are două alele AA în locusul luat în considerare, ea trebuia să fi primit un A de la fiecare părinte; astfel vedem că tatăl și-a transmis alela A și nu G și mama i-a transmis și alela A și nu G, indicând că aceste două transmisii aparțin celulei b care descrie transmiterea unei alele A și nu a unei G.

AG **AG**

AA

	Transmitted	*Allele*
Not transmitted	Allele G	Allele A
Allele G	Cell a	Cell b = 2
Allele A	Cell c	Cell d

Trio 2: Aici o transmisie a fost a unui A și nu a unui G, iar cealaltă a fost a unui G și nu a unui A. (Nu știm care a fost de la ce părinte, dar știm că 1 transmisie este descrisă de celula b și cealaltă de celula c.)

		Transmitted	*Allele*
AG	**AG**	*Not transmitted*	Allele G	Allele A
AG		Not G		1 (mother **or** father)
		Not A	1 (mother **or** father)	

Trio 3: Aici, A din copil trebuia să vină de la tată, deoarece asta era tot ce trebuia să transmită, așa că a transmis o alele A și nici nu și-a transmis cealaltă alele A; astfel că transmisia aparține celulei d. Mama și-a transmis alela G și nu alela A și astfel că transmiterea aparține celulei c.

		Transmitted	*Allele*
AA	**AG**	*Not transmitted*	Allele G	Allele A
AG		Not G		
		Not A	1 (mother)	1 (father)

Imaginează-ți că avem 120 de astfel de triouri. Acum vom combina datele din cele 240 de transmisii dintre cele 120 de triouri adăugând numerele din fiecare celulă într-un tabel rezumativ, așa cum se arată în exemplul de mai jos:

	Transmis	*Alela*
Netransmis	Alele G	Alela A
Nu G	65	90
Nu A	50	45

și calculați statistica de testare TDT ca cantitate:

$$\frac{(b - c)^2}{(b + c)} \qquad \frac{(90 - 50)2}{140} = 11.43$$

care este distribuit ca Z^2, cu un grad de libertate, și deoarece este mai mare de 3,84, putem respinge ipoteza lipsei de legături și putem concluziona că există dovezi de legătură și asociere. TDT este într-adevăr un test McNemar (descris în secțiunea 3.2), analog analizei caz-control potrivite, în care controalele sunt alele netransmise, mai degrabă decât persoane.

Rețineți că orice transmisii care ajung în celula a sau d sunt neinformative. Statistica de testare TDT utilizează numai informații din celulele b și c. Astfel, doar părinții care sunt heterozigoți (având un A și un G) sunt informativi. Din moment ce începem cu copiii afectați identificați prin fenotip, nu știm dacă părinții sunt heterozigoți și trebuie să genotipeze toți părinții din colecția noastră de triouri, deși unii se vor dovedi a nu fi informativi. De asemenea, în exemplele noastre de triouri, copiii afectați aveau fiecare cel puțin o alelă A, dar este posibil ca copilul să fie afectat și să aibă un genotip GG, deoarece pot exista și alte gene care conferă susceptibilitate bolii.

Mărimea eșantionului necesară pentru o putere suficientă pentru a detecta legătura și asociere prin TDT depinde de mulți factori, inclusiv frecvența markerului și al alelei bolii, fracția de recombinare sau dezechilibrul de legătură între marker și alela bolii și mărimea efectului alelei bolii. Un statistician genetic este cel mai capabil să calculeze dimensiunea eșantionului și ar trebui să fie consultat înainte de a întreprinde orice astfel de muncă. [34]

TDT a fost extins pentru a fi aplicabil în cazurile în care ADN-ul unuia sau ambilor părinți nu este disponibil, la frați și la alte grupuri de familie, precum și la trăsăturile cantitative.

Pentru a fi mai concret, imaginați-vă un studiu al geneticii schizofreniei. Să presupunem că studiile anterioare au legat schizofrenia de o regiune de pe cromozomul 22 și că există o genă în această regiune pe care o considerăm o genă candidată puternică, deoarece este implicată în producerea unui neurotransmițător despre care se crede că este implicat în boală . Mai știm că există un SNP în această genă care afectează secvența

de aminoacizi a proteinei produse de genă. Vrem să testăm dacă acest SNP este asociat cu schizofrenia. Avem un eșantion de 200 de triouri formați dintr-un urmaș cu schizofrenie și ambii părinți. Fiecare părinte poate transmite descendenților afectați una dintre cele două alele ale sale la acest locus. Pentru fiecare trio, putem construi un tabel 2x2 și numărăm numărul de alele transmise care aparțin fiecărei celulă a tabelului și apoi combinam tabelele pentru a obține un tabel rezumat pe care îl folosim pentru a calcula statistica TDT așa cum se arată mai sus. Dacă această statistică este mai mare de 3,84, putem respinge ipoteza lipsei unei legături la nivelul semnificației de .05.

8.8 Câteva concepte și complexități suplimentare ale studiilor genetice

Cum selectăm *genele candidate?* O strategie este de a analiza calea biologică implicată în boală, de a lua în considerare o proteină relevantă în acea cale și de a explora polimorfismele genelor legate de acea proteină. O altă posibilitate este să ne uităm la genele dintr-o regiune de pe cromo care este legată de boală, chiar dacă nu știm ce fac aceste gene de fapt. Studiile de asociere analizează genele candidate, dar nu avem candidați convingătoare pentru multe boli. Tehnologia este disponibilă pentru a face *scanări de asociere a întregului genom, folosind metode de asociere pentru a testa alelele prin examinarea* polimorfismelor din întregul genom. Cu toate acestea, acoperirea genomului în acest mod poate necesita testarea multor mii de SNP, creând o problemă de testare a ipotezei multiple și o probabilitate mare de eroare de tip I. (A se vedea secțiunea 3.24.)

Analizele de haplotip și *hărțile de haplotip* sunt căi promițătoare pentru a reduce numărul de polimorfisme care trebuie examinate pentru a găsi gene de susceptibilitate la boli. Haplotipurile sunt secvențe de alele de-a lungul unui cromozom dat. Conceptul de dezechilibru de legătură ne spune că alelele care sunt apropiate din punct de vedere fizic pot fi moștenite împreună. Aceste haplotipuri cu dezechilibru de legături ridicate, sau „blocuri haplotip", pot fi conservate de-a lungul generațiilor, deoarece sunt moștenite ca „bloc", iar studiile recente au arătat că există o diversitate limitată a haplotipurilor în populația umană. Se pare că există mai multe alele în cadrul unor astfel de haplotipuri care caracterizează întregul

haplotip și ar putea fi posibil să ne uităm doar la acele alele cheie în loc de fiecare polimorfism de nucleotidă separat.

Printre lucrurile care complică studiile genetice se numără următoarele:

Este adesea dificil să identifici *un fenotip* suficient de specific. Să presupunem că căutăm gena pentru hipertensiune — cum definim o - hipertensiune când tensiunea arterială este o variabilă continuă? Hipertensiunea arterială a fost definită ca fiind peste un anumit punct de tăiere al tensiunii arteriale pe baza predicțiilor riscurilor de morbiditate și mortalitate asociate cu diferite niveluri, dar care nu corespunde neapărat unor caracteristici ale eredității.

Sau luați gene care predispun la boli de inimă. Trebuie să definiți boala cardiacă foarte specific - include insuficiența cardiacă (care poate fi un proces diferit de boală); include doar atacurile de cord precoce? În psihiatrie, diagnosticele se bazează adesea pe un număr „x" de simptome dintr-un posibil „n" simptome, dar cum rămâne cu persoanele care au „x - 1" astfel de simptome? Categoria de diagnosticare poate să nu fie cea ereditară. O abordare a acestor probleme de definire a fenotipului este identificarea componentelor de bază sau a fenotipurilor intermediare care pot fi o expresie mai directă a efectelor genelor decât sunt diagnosticele complexe utilizate în medicină. Exemple de astfel de fenotipuri intermediare includ nivelurile de IgE în studiile genetice ale astmului, imagistica cerebrală funcțională în studiile genetice ale schizofreniei și nivelurile de lipoproteine în studiile bolilor cardiace aterosclerotice.

Majoritatea bolilor au moduri complexe de moștenire. Nu există o genă care determină susceptibilitatea, ci mai multe, fiecare dintre ele contribuind o cantitate modestă și care pot interacționa între ele. Pătrunderea genelor variază . Aceasta înseamnă că, chiar dacă aveți alela bolii, este posibil să nu aveți boala, fie pentru că sunt implicate și alte gene, fie pentru că factorii de mediu specifici sunt o condiție necesară pentru ca gena să funcționeze. Foarte puține gene sunt complet penetrante; Boala Huntington este una care este: persoanele care au gena bolii Huntington vor face boala.

S-ar putea să vă uitați la fenotipuri care sunt într-adevăr *fenocopii* . Fenocopiile sunt fenotipuri ale bolii care nu sunt cauzate de factori genetici, ci mai degrabă de factori de mediu. Includerea din neatenție a fenocopiilor în analizele dumneavoastră va acționa ca zgomot și va dilua orice descoperire legată de influențele genetice adevărate.

Polimorfismele din diferite gene pot produce fenotipuri foarte

asemănătoare. Acest lucru este cunoscut sub numele de *eterogenitate genetică* . Alte complicații apar din *epistasis* care apare atunci când expresia genei bolii este afectată de acțiunea unei alte gene. Deci, după cum puteți vedea, domeniul este foarte complex și statisticile pentru a evalua dacă constatările legate de legături și asocieri sunt mai mult decât întâmplătoare sunt, de asemenea, complexe. [36,37,38] Noi tehnologii (cum ar fi scanarea asocierii întregului genom) și noi instrumente analitice (cum ar fi „hapmaps") sunt dezvoltate rapid pentru a rezolva astfel de probleme.

A existat o schimbare de paradigmă în știință – de la a crede că lucrurile sunt mai simple decât par la a înțelege că sunt mai complexe decât par. În ultimul secol, principiul care a călăuzit eforturile științifice a fost briciul lui Occam - că explicația cea mai parcimonioasă pentru fenomene este cea mai bună. Dar, pe măsură ce descoperirile genomice și moleculare se accelerează, devine evident că în sfera biologică explicațiile simple nu sunt posibile și scopul este de a descoperi și explica mai precis complexitatea (și minunea) inerentă a vieții.

Capitolul 9
ETICA ȘI STATISTICA CERCETĂRII

Moralitatea, ca și arta, înseamnă să tragem o linie undeva.

Oscar Wilde (1854-1900)

9.1 Ce legătură are statisticile cu asta?

La prima vedere poate părea că statistica și etica cercetării nu au nimic de-a face una cu cealaltă. Nu așa! Luați în considerare de ce atât de mulți oameni se oferă voluntari pentru studii de cercetare medicală. În multe cazuri, se datorează faptului că există un beneficiu așteptat. De exemplu, în studiile clinice pentru cancer, deseori, medicamentul experimental este o ultimă speranță și este posibil să nu fie disponibil în afara studiului. În multe studii privind bolile cardiovasculare, participanții apreciază îngrijirea și atenția suplimentară și sunt dispuși să încerce un nou medicament, de exemplu, pentru hipertensiune arterială. Și, de fapt, s-a demonstrat că din zece participanții la studiile clinice trăiesc mai mult și se descurcă mai bine decât populația generală, chiar dacă sunt tratați cu un placebo. Dar ceea ce poate nu este suficient de apreciat este faptul că mulți, mulți oameni participă la studii din altruism pentru a avansa cunoștințele științifice. Cunoștințele științifice nu sunt avansate atunci când un studiu este prost conceput, sau realizat fără suficientă rigoare sau nu suficient de mare pentru a da un răspuns. Statisticile adecvate sunt un factor determinant al eticii unui studiu.

Un prim exemplu este Inițiativa pentru sănătatea femeilor (WHI), descrisă în capitolul 6. Femeilor aflate în postmenopauză li sa cerut să se alăture unui studiu de terapie de substituție hormonală; studiul ar continua până la 12 ani înainte ca rezultatele să fie cunoscute și ar putea să nu beneficieze în mod direct femeile în sine, dar ar răspunde la întrebarea importantă a efectului hormonilor asupra cancerului, bolilor de inimă și osteoporozei . Mulți dintre participanții WHI au participat pentru fiicele și nepoatele lor și și-au exprimat mândria și entuziasmul pentru răspunsul la întrebările pentru generațiile viitoare. Și într-adevăr au atins acest obiectiv – o parte din WHI, studiul estrogen plus progestativ versus placebo, a

răspuns deja la aceste întrebări importante, cu un

, crește atacurile de cord, accidentul vascular cerebral și cheaguri de sânge și demența. Deci, deși tratamentul arată beneficii în ceea ce privește cancerul colorectal și fracturile osteoporotice, riscurile generale depășesc beneficiile . Acest proces a schimbat practica medicală pentru generațiile viitoare.

Ei bine, asta ne readuce la statistici. Acest studiu a putut să răspundă la aceste întrebări deoarece a avut suficientă putere pentru a le răspunde. A fost nevoie de 16.608 de femei din acea parte a WHI pentru a putea detecta aceste efecte . Chiar dacă rezultatul ar fi fost nul (adică dacă nu a arătat nicio diferență între grupurile de tratament și placebo), am fi putut avea încredere în acel rezultat, deoarece puterea era acolo de a detecta un efect adevărat dacă exista într-adevăr unul. După cum sa dovedit, rezultatele au fost clare, deși neașteptate, în favoarea placebo. Deci, ideea este că, pentru ca un studiu să fie „etic", trebuie să fie proiectat și alimentat suficient de bine, astfel încât să poată răspunde la întrebările pe care le pune. În caz contrar, este posibil ca persoanele care acceptă să participe în așteptarea că vor contribui la cunoaștere, de fapt, să nu contribuie, deoarece studiul este prost conceput, alimentat sau executat și pot fi expuși inutil la riscuri.

Rețineți că există anumite modele de studiu pentru care considerațiile de putere sunt mai puțin relevante. Exemple sunt studiile pilot, care, prin definiție, sunt menite să testeze fezabilitatea unui protocol de cercetare sau să culeagă date preliminare pentru a planifica un studiu complet, sunt scutite de problema puterii. Considerațiile de putere s-ar putea să nu se aplice, de asemenea, anumitor studii de toxicitate a medicamentelor (studii de fază I) sau anumitor tipuri de studii de cancer, dar cu siguranță în studiile de prevenire, precum și în studiile de tratament de fază III, puterea este o considerație majoră în etica cercetării.

9.2 Protecția subiectelor de cercetare umane

Subiecții umani din cercetarea medicală contribuie foarte mult la îmbunătățirea sănătății oamenilor. Acești voluntari trebuie să fie protejați de vătămări cât mai mult posibil. În trecutul nu prea îndepărtat, au existat unele încălcări flagrante ale principiilor etice în efectuarea cercetării medicale. Cele mai îngrozitoare exemple din lume sunt experimentele medicale efectuate în lagărele de concentrare naziste de către medici! Sfidează orice

un fel de înțelegere a modului în care profesioniștii educați, probabil „civilizați” și-ar fi putut denatura profesia și propria lor umanitate . Dar aceste atrocități au avut loc și au demonstrat ororile pe care oamenii sunt capabili să le comită. Când aceste atrocități au devenit cunoscute după cel de-al Doilea Război Mondial, procesele de la Nuremberg ale criminalilor de război naziști (inclusiv medicii care au efectuat astfel de cercetări) au dus și la Codul de la Nürnberg pentru desfășurarea cercetării medicale, care a stabilit cerința de bază a consimțământului informat voluntar. Ulterior, Declarația de la Helsinki, din 1964, a extins și rafinat liniile directoare de cercetare și a devenit un standard mondial, care suferă revizuiri periodice .

Cel mai infam exemplu de cercetare lipsită de etică din SUA a fost probabil studiul Tuskegee Institute despre sifilis, care a avut loc în sudul Statelor Unite între 1932 și 1972. Cercetătorii au vrut să studieze cursul natural al sifilisului. În anii 1940 au devenit disponibile antibiotice care puteau trata această boală, dar au fost refuzate participanților, care erau bărbați săraci de culoare, deoarece o intervenție pentru tratarea bolii ar interfera cu acest studiu observațional. În 1972 publicul a luat cunoștință de acest experiment și în 1974 a fost înființată Comisia Națională pentru Protecția Subiecților Umani de Cercetare Biomedicală și Comportamentală. Ei au elaborat un raport cunoscut sub numele de *Raportul Belmont:* principii etice și linii directoare pentru protecția subiecților umani de cercetare. Aceste linii directoare sunt urmate de toate școlile de medicină și alte instituții de cercetare care efectuează cercetări care implică participanți umani și sunt considerate a fi principii universale, care traversează liniile culturale. Orientările se bazează pe trei principii de bază: *respectul pentru persoane, binefacerea și justiția.*

Respectul pentru persoane recunoaște că oamenii sunt ființe autonome și pot face propriile lor alegeri informate cu privire la participarea la - cercetare, fără constrângere. Procesul de consimțământ informat se bazează pe acest principiu. Participanții care nu sunt capabili să facă propriile alegeri, cum ar fi pacienții în comat sau persoanele cu incapacități mintale sau copiii mici, trebuie să aibă protecții speciale.

Beneficiul, sau principiul non-afectiunii, înseamnă că riscurile cercetării trebuie reduse la minimum, beneficiile maximizate, iar cercetătorul este responsabil pentru protejarea participantului.

Justiția în acest context se referă la o distribuție echitabilă a riscurilor și beneficiilor cercetării. Un grup de persoane nu ar trebui să fie expus - riscurilor de cercetare în beneficiul altui grup de persoane. Acesta poate

deveni un concept destul de complicat. Deși poate fi ușor să discerneți încălcări în anumite situații - pentru a lua exemplul cel mai extrem, prizonierii naziștilor au fost supuși unor experimente de înghețare în beneficiul soldaților care ar putea fi nevoiți să lupte în condiții arctice - poate fi mai subtil în multe situații și acestea trebuie examinate cu atenție, conform acestui principiu.

9.3 Consimțământ informat

Unul dintre cele mai importante elemente în protecția subiecților umani este principiul consimțământului informat. Subiectul studiului trebuie să - consimtă în mod liber să facă parte din studiu după ce a fost pe deplin informat cu privire la riscurile și beneficiile potențiale.

Există anumite elemente care trebuie să fie într-o formă scrisă de consimțământ. Trebuie precizat scopul cercetării; trebuie listată o persoană de contact de 24 de ore; trebuie să existe o descriere a procedurilor de studiu: ce se așteaptă de la participant, durata studiului și cât va dura timpul participantului. Potențialele riscuri și neplăceri , potențiale beneficii, inconveniente pentru participanți, toate trebuie să fie clar precizate. Trebuie să existe o declarație conform căreia participarea este voluntară și că participantul are dreptul de a se retrage în orice moment și că acest lucru nu va prejudicia grija participantului. În cazul în care cercetarea poate duce la nevoie de îngrijire suplimentară sau proceduri de diagnosticare, participantului trebuie să i se spună în ce măsură este responsabil pentru îngrijirea ulterioară și pentru ce va plăti studiul. Dacă există vreo compensație pentru participanți, fie pentru cheltuielile efectuate în timpul participării, fie pentru timpul petrecut, aceștia trebuie să fie informați cu privire la sumă. (Suma nu trebuie să fie excesivă , deoarece poate părea coercitivă.) Trebuie inclusă o declarație care să asigure confidențialitatea și modul în care va fi păstrată.

Cel mai important, participantul trebuie să înțeleagă cu ce este de acord și formularul de consimțământ trebuie să fie redactat într-un limbaj care este ușor de înțeles și, dacă este cazul, tradus în limba maternă a participantului. Toate acestea trebuie să fie aprobate de IRB (sau Institutional Review Board) al instituției medicale, care este în general un comitet de experți și neprofesionați care revizuiesc și trebuie să aprobe toate protocolurile de cercetare înainte de începerea cercetării și care monitorizează evenimentele

adverse pe măsură ce cercetarea progresează. Diferitele IRB-uri au cerințe specifice diferite care sunt de obicei postate pe site-urile lor web. Consimțământul informat este un proces continuu – nu este doar semnarea unui formular la începutul unui studiu. Cercetatorul are obligația de a informa participantul cu privire la noile cercetări relevante care i-ar putea afecta decizia de a continua să participe.

Înapoi la WHI - deoarece se credea la momentul în care a fost început WHI că hormonii vor proteja femeile de boli de inimă, formularul inițial de consimțământ a precizat acest lucru ca un potențial beneficiu. Riscurile potențiale menționate în formularul de consimțământ au inclus o creștere a cancerului de sân și a cheagurilor de sânge. Când WHI era în desfășurare, HERS (Heart and Estro gen Replacement Study) a publicat rezultate care indică faptul că pentru femeile care aveau deja boli de inimă (proces de prevenire secundară), înlocuirea hormonală nu a oferit niciun beneficiu. Ei au observat mai multe atacuri de cord în partea incipientă a studiului, cu o posibilă reducere târzie, ducând la nicio diferență generală între grupurile de tratament și placebo până la sfârșitul studiului. Aceste informații au fost furnizate printr-o corespondență specială tuturor femeilor care participă la programul hormonal WHI pentru prevenirea primară a bolilor de inimă. (Prevenția primară înseamnă că studiul a fost efectuat pe femei în general sănătoase). Ulterior, datele timpurii de la WHI în sine au indicat că a existat un prejudiciu precoce în ceea ce privește bolile de inimă . Din nou, femeile au fost informate printr-o corespondență specială, telefon și discuții personale cu personalul clinicii. În cele din urmă, studiul cu estrogen plus progestativ a fost oprit după 5,2 ani (în loc de media planificată inițial de 8,5 ani), deoarece excesul de risc de cancer de sân a depășit o limită de oprire predeterminată și un indice global al efectelor generale a sugerat mai mult rău decât beneficiu, iar toate femeile din studiu au fost întrerupte din pastilele de studiu.

9.4 Echilibru

Asta ne aduce la un alt concept: când este etic să începem un studiu clinic al unui nou tratament? Când există echilibru. *Echilibrul* înseamnă că există dovezi aproximativ egale că tratamentul poate oferi beneficii și că nu va oferi beneficii. Dacă suntem siguri că tratamentul oferă beneficii, nu ar trebui să-l refuzăm persoanelor care ar primi placebo în cadrul studiului. Bineînțeles că putem greși. Au existat critici la adresa Women's Health Initiative, care au spus că nu este etică să faci un astfel de studiu, deoarece

era bine cunoscut faptul că hormonii protejează împotriva bolilor de inimă și ar fi lipsit de etică să refuzi acești hormoni femeilor randomizate la placebo! Bineînțeles că acum știm că a fost greșit - grupul placebo s-a descurcat mai bine. La momentul începerii WHI, dovezile observaționale indicau beneficii în ceea ce privește bolile de inimă, dar nu fusese niciodată testată într-un studiu clinic, care este „standardul de aur". Astfel, au fost mulți oameni care nu au crezut că beneficiile înlocuirii hormonale erau deja stabilite de studiile observaționale și se dovedește că aveau dreptate. Cercetătorul, a cărui obligație este să protejeze participanții la cercetare umană, trebuie să creadă că este la fel de probabil că tratamentul este mai bun sau că placebo sau tratamentul de comparație este mai bun. Comunitatea științifică care judecă propunerea de cercetare trebuie să creadă, pe baza „state-of-the-art", că există o întrebare rezonabilă la care trebuie să se răspundă.

9.5 Integritatea cercetării

Pentru ca concluziile cercetării să fie valide, procedurile de colectare a datelor trebuie să fie administrate riguros și uniform. Nicio dată nu poate fi modificată fără documente. Dacă există o eroare materială, modificarea și motivul acesteia trebuie documentate. Înscrierea trebuie să fie conform standardelor stricte și pre-planificate. Uneori (din fericire, rar) există o presiune mare pentru înscrierea subiecților într-un interval de timp dat, sau cercetătorul (cu încălcarea principiului echilibrului) crede cu adevărat că tratamentul îi poate ajuta pe pacienții săi și astfel „îndoiește" regulile de înscriere. Acest lucru poate invalida cercetarea și, prin urmare, nu este etică. Un exemplu foarte trist în acest sens a avut loc în Proiectul Național Chirurgical Adjuvant Breast and Bowel (NSABP). Acest studiu multicentric a demonstrat că lumpectomia ar putea fi echivalentă cu mastectomia la sute de mii de femei. Președintele acestui studiu a descoperit că investigatorul principal dintr-unul din centrele clinice a falsificat unele fișe ale pacienților, astfel încât femeile care nu erau eligibile pentru a participa la studiu pe baza unor criterii de înscriere predeterminate, au fost făcute fals eligibile pentru a participa. Acest studiu excelent și extrem de important a fost inițial afectat atunci când a devenit cunoscut și președintele studiului a fost acuzat de Oficiul de Integritate a Cercetării (ORI) cu o - conduită științifică greșită, chiar dacă a notificat NIH cu privire la problemă

când a aflat despre aceasta. Ulterior, a fost eliberat complet și i s-au cerut scuze multiple. Studiul a avut implicații profunde asupra tratamentului femeilor cu cancer de sân. Cu toate acestea, aceasta a fost o încălcare gravă a eticii din partea unui anchetator dintr-unul dintre numeroasele centre, care ar fi putut invalida constatările. Din fericire, rezultatele au durat chiar și atunci când au fost excluși toți pacienții din clinica infracțională.

9.6 Politicile de autor

În cercetarea medicală, majoritatea articolelor de cercetare originale au mai mulți autori, deoarece cercetarea medicală este un efort de colaborare. Majoritatea revistelor medicale și instituțiilor de cercetare au politici specifice și stricte de autor (publicate în reviste și/sau pe site-uri web), multe dintre ele cuprind următoarele elemente: (1) coautorii trebuie să aducă o contribuție intelectuală la lucrare (de exemplu, să conceapă cercetarea, să efectueze analize, să scrie secțiuni ale lucrării sau să facă contribuții editoriale); (2) toți co- autorii trebuie să poarte responsabilitatea pentru conținutul acestuia; (3) Co-autorii trebuie să dezvăluie potențialele conflicte de interese (de exemplu, sprijin relevant din industrie, lectorate, deținerea de acțiuni). Ordinea calității de autor poate fi uneori un punct de dispută și ar trebui discutată de co- autorii la începutul procesului.

9.7 Comitete de monitorizare a datelor și siguranței

În general, studiile clinice au un comitet de monitorizare a datelor și siguranței (DSMB) pentru a supraveghea studiul. Acestea sunt grupuri independente de experți în disciplinele relevante, care au o calitate consultativă. Sarcina lor este să monitorizeze procesul și să asigure siguranța participanților. Într-o încercare în orb, ei sunt singurii care văd datele neoarbe la intervale regulate, prestabilite. Dacă găsesc beneficii sau vătămări excesive într-un braț al studiului, ei ar sfătui să oprească studiul (așa cum sa întâmplat în Inițiativa pentru sănătatea femeilor). De obicei, criteriile pentru oprirea unui studiu din cauza vătămării în grupul de tratament sunt mai stricte decât oprirea pentru beneficii.

9.8 Rezumat

Conducerea etică a cercetării are multe componente. Apar întrebări etice noi și dificile pe măsură ce știința progresează și noile tehnologii devin disponibile. Acest scurt capitol începe să vă ofere o idee despre unele dintre problemele implicate. Informații mult mai detaliate sunt disponibile de pe diverse site-uri web, iar NIH are un curs on-line pentru protecția subiecților umani. IRB-urile locale vă pot oferi informații și surse suplimentare.

Post-scriptum
CÂTEVA COMENTARII DE DESPĂZIRE ASUPRA IMPACTULUI EPIDEMIOLOGIEI ASUPRA VIEȚILOR OMULUI

În urmă cu zece ani, o femeie cu cancer de sân ar fi probabil să facă o mastectomie radicală, care, pe lângă îndepărtarea sânului și desfigurarea rezultată, ar include și îndepărtarea unei mari părți a peretelui muscular din piept și ar lăsa-o incapacită în multe feluri. Astăzi , aproape nimeni nu face o mastectomie radicală și mulți nici măcar nu fac o mastectomie modificată, dar, în funcție de cancer, poate face o pectomie de lum, care doar îndepărtează nodul, lăsând sânul intact. Cu ani în urmă, nimeni nu a acordat prea multă atenție radonului, un gaz inert eliberat din sol și disipat prin crăpăturile fundației în case. Acum este recunoscut ca principala cauză a cancerului pulmonar. Rolul nutriției în prevenirea bolilor nu a fost recunoscut de comunitatea științifică. De fapt, oamenii care credeau în importanța nutrienților în cauza și vindecarea bolii erau considerați a fi niște înfățișați, doar puțin nebuni. Acum este adesea subiect de articole, cărți și știri, iar sume substanțiale de fonduri de cercetare sunt investite în studii nutriționale. Astfel de studii influențează legislația, cum ar fi, de exemplu, reglementările conform cărora alimentele procesate trebuie să aibă etichetare standard, ușor de înțeles de publicul larg, a conținutului de grăsimi din alimente, precum și de sodiu, vitamine și alți nutrienți. Toate acestea au un impact asupra schimbării obiceiurilor alimentare ale populației, precum și asupra economiei industriei alimentare.

În domeniul sănătății, schimbările în tratamentul, prevenirea și cunoștințele predominante apar atunci când există o confluență de circumstanțe: se dobândesc informații noi pentru a înlocui teoriile existente; există o diseminare a acestor informații către comunitatea științifică și către publicul larg; și există climatul psihologic, economic și politic adecvat care ar saluta adoptarea noilor abordări . Epidemiologia joacă un rol major prin furnizarea metodelor prin care se dobândesc noile cunoștințe științifice. Adesea, primele indicii asupra cauzalității vin cu mult înainte ca un mecanism biologic să fie cunoscut. În jurul anului 1850, la Londra, dr. John

Snow, consternat de suferința și decesele cauzate de epidemiile de holeră, a studiat cu atenție rapoartele despre astfel de epidemii și a remarcat că holera era mult mai probabil să apară în anumite părți ale Londrei decât în alte părți. El a cartografiat locurile în care holera era răspândită și unde era mai puțin și a remarcat că casele furnizate cu apă de o companie, Southwark and Vauxhall Company, aveau mult mai multe cazuri de holeră decât cele furnizate de o altă companie. De asemenea, știa că compania Vauxhall folosea ca sursă o zonă puternic contaminată de canalizare. Zăpada a impus ca orașul să închidă pompa care furnizează apă contaminată, cunoscută sub numele de Broad Street Pump. Au făcut așa și holera s-a diminuat. Toate acestea au fost cu 25 de ani înainte ca cineva să izoleze bacilul holeric și cu mult înainte ca oamenii să accepte ideea că boala poate fi răspândită prin apă. În vremurile moderne, epidemia de SIDA este una în care metoda de răspândire a fost identificată înainte ca agentul infecțios, virusul HIV, să fie cunoscut.

Tehnicile epidemiologice au fost din ce în ce mai aplicate bolilor cronice, care diferă de bolile infecțioase prin faptul că pot persista o perioadă lungă de timp (în timp ce infecțiile de obicei fie ucid rapid, fie se vindecă rapid) și, de asemenea, au de obicei cauze multiple, dintre care multe sunt dificil de identificat. Și aici, epidemiologia joacă un rol central în identificarea factorilor de risc, cum ar fi fumatul pentru cancerul pulmonar. O astfel de cunoaștere este transpusă în acțiune publică înainte ca toate căile biologice să fie elucidate. Acțiunea ia forma unor campanii educaționale, legi împotriva fumatului, restricții privind reclamele și alte mecanisme de limitare a fumatului. Factorii de risc pentru bolile de inimă au fost identificați prin studii epidemiologice clasice care au dus la schimbări ale stilului de viață pentru indivizi, precum și consecințe de politică publică.

Bolile cronice prezintă probleme diferite și provocatoare în analiză, iar noi tehnici statistice continuă să fie dezvoltate pentru a se adapta acestor probleme. De asemenea, sunt dezvoltate noi tehnici statistice pentru problemele speciale întâlnite în cercetarea genetică. Astfel, domeniul statisticii nu este static, iar domeniul epidemiologiei nu este fix. Ambele se adaptează și se extind pentru a face față problemelor de sănătate în schimbare ale societății noastre.

Anexa A
VALORI CRITICE ALE CHI-PĂTRATUL, Z ȘI t

Când valoarea Z, Z2 sau t calculată din datele observate este egală cu sau depășește valoarea critică enumerată mai jos, putem respinge ipoteza nulă la nivelul de semnificație dat, ^{a} *(* alfa).

Valori critice selectate ale Chi-pătratului

Nivel de semnificație	.1	.05	.01	.001
Valoarea critică a / ^{2}	2,71	3,84	6,63	10.83

Valorile critice selectate ale lui Z

Nivel de semnificație Test cu două cozi (Test cu o coadă)	.1 (.05)	.05 (.025)	.01 (.005)	.001 (.0005)
Valoarea critică a lui	1,64	1,96	2,58	3.29

Valorile critice selectate ale t

Nivel de semnificație Test cu două cozi (Test cu o coadă)	.10 (.05)	.05 (.025)	.01 (.005)	.001 (.0005)
Grade de libertate				
9	1,83	2.26	3.25	4,78
19	1,73	2.09	3,86	3,88
100	1,66	1,98	2,63	3.39
1000	1,64	1,96	2,58	3.29

NOTĂ: Interpretare:
Dacă aveți 19 grade de libertate, pentru a respinge H_o, la a = .05 cu un test cu două cozi, ați avea nevoie de o valoare a t la fel de mare sau mai mare de 2,09; pentru a = .01, ar fi nevoie de cel puțin la fel de mare ca 3,86. Rețineți că atunci când df devine foarte mare, valorile critice ale lui t sunt aceleași cu valorile critice ale lui Z. Alte valori decât cele calculate aici apar în majoritatea textelor prezentate în lecturile sugerate.

Anexa B
TESTUL EXACT AL FISHER

Să presupunem că aveți un tabel 2 pe 2 care rezultă dintr-un experiment pe șobolani care expune un grup la o anumită condiție experimentală și celălalt grup la o condiție de control, cu rezultatul de a fi în viață după o săptămână. Tabelul arată după cum urmează:

Tabelul B.1

	Controla	Experimental	
În viaţă	*o* 1	*b* 7	Rândul 1 = R_1 = 8
Mort	*c* 5	*d* 1	Rândul 2 = R_2 = 6
Total	Col 1 = C_1 = 6	Col 2 = C_2 =	*N* = 14

87,5% din grupul experimental și 16,7% din grupul de control au trăit. Un rezultat mai extrem, având în vedere aceleași totaluri de rând și coloană, ar fi

Tabelul B.2

	Controla	Experimental	
În viaţă	0	8	8
Mort	6	0	6
Total	6	8	14

unde au trăit 100% din grupul experimental și 0% din grupul de control. Un alt rezultat mai extrem ar fi acolo unde au trăit 25% dintre cei experimentali și toți martorii:

Tabelul B.3

	Controla	Experimental	
În [illegible]	6	2	8
Mort	0	6	6
Total	6	8	14

(Orice alte tabele pe care le-am putea construi cu aceleași totaluri marginale ar fi mai puțin extreme decât tabelul B.1, deoarece nicio celulă nu ar conține un număr mai mic decât cel mai mic număr din tabelul B.1, care este 1.)

Calculăm probabilitatea exactă de a obține rezultatul observat al experimentului doar întâmplător (Tabelul B.1), sau unul chiar mai extrem (ca în Tabelul B.2 sau B.3), dacă ar fi cu adevărat adevărat că nu au existat diferențe de supraviețuire între cele două grupuri. Testul exact al lui Fisher se calculează obținând probabilitatea fiecăruia dintre aceste tabele și însumând aceste probabilități.

Mai întâi trebuie să explicăm simbolul Se numește „factorial". Un număr *n!* înseamnă (*n)* x (*n* -1) x (*n* -2) x . . . x (1). De exemplu, 6! = 6 x 5 x 4 x 3 x 2 x 1. Prin definiție, 0! este egal cu 1.

Probabilitatea de a obține observațiile din tabelul B.1 este

$$\frac{(*_1!) \times (\wedge!) \times (C_1!) \times (C_2!\wedge 8!<5!6!8!}{a! \times b! \times c! \times d! \times N!} \quad {}_0 15_{984} \qquad 1!7!5!1!14!\ '$$

Probabilitatea de a obține observațiile din tabelul B.2 este

$$\frac{8!6!6!8!}{0!8!6!0!14!} = .000333$$

Probabilitatea de a obține observațiile din tabelul B.3 este

$$\frac{8!6!6!8!}{6!2!0!6!14!} = .009324$$

Suma acestor probabilități este .015984 + .000333 + .009324 = .025641. Astfel putem spune că probabilitatea exactă de a obține rezultatele pe care le-am observat în tabelul B.1, sau rezultate mai extreme, este .025641, dacă ipoteza nulă a fost adevărată. Putem respinge ipoteza nulă că rata de supraviețuire este aceeași în ambele grupuri la un nivel de semnificație p = .026.

Anexa C
TEST NEPARAMETRIC KRUSKAL-WALLIS PENTRU COMPARAREA MAI MULTE GRUPE

De exemplu, să presupunem că aveți trei grupuri de oameni fiecare având un scor pe o anumită scară. Numărul total de persoane din toate cele trei grupe este *N*. Procedura generală este următoarea: (1) Combinați toate scorurile din cele trei grupe și ordonați-le de la cel mai mic la cel mai mare. (2) Se acordă rangul de 1 celui mai mic punctaj, 2 următorului cel mai mic punctaj și așa mai departe, *N* fiind atribuit persoanei cu cel mai mare punctaj. (3) Sortați oamenii înapoi în grupurile lor originale, fiecare persoană având rangul său atribuit. (4) Însumați toate rangurile din fiecare grupă. (5) Calculați cantitatea prezentată mai jos, pe care o numim *H*. (6) Dacă aveți mai mult de cinci cazuri în fiecare grup, puteți căuta *H* într-un tabel chi-pătrat, cu *k* -1 grad de libertate (unde *k* este numărul de grupuri comparate).

Scoruri la înțelegerea lecturii					
Grupa A		Grupa B		Grupa C	
Scoruri	(Rang)	Scoruri	(Rang)	Scoruri	(Rang)
98	(13)	80	(9)	120	(21)
70	(6)	60	(2)	110	(17)
68	(5)	106	(15)	90	(12)
107	(16)	50	(1)	114	(19)
115	(20)	75	(8)	105	(14)
65	(4)	74	(7)	85	(10)
Sumă	**(64)**	64	(3)	112	(18)
de			**(45)**	87	(11)
Ranguri					**(122)**

$$\frac{12}{N(N+1)}\left[\frac{64^2}{6}+\frac{45^2}{7}+\frac{122^2}{8}\right]-3(N+1)$$

$$= 7{,}57$$

grade de libertate = 3-1 = 2

Dacă ipoteza nulă a lipsei diferențelor de rang mediu între grupuri a fost adevărată, probabilitatea de a obține un chi-pătrat la fel de mare sau mai mare decât 7,57 cu 2 grade de libertate este mai mică de .05, așa că putem respinge ipoteza nulă și putem concluziona că grupurile diferă. (Când apar egalități în clasament, fiecărui scor i se dă media rangului pentru care este egal. Dacă există multe egalități, poate fi utilizată o corecție la *H* , așa cum este descris în cartea lui Siegel enumerată în Lecturi sugerate.)

Anexa D
CUM SE CALCULEAZĂ UN COEFICIENT DE CORELARE

Individual	*X*	*Y*	X^2	Y^2	*XY*
O	5	7	25	49	35
B	8	4	64	16	32
C	15	8	225	64	120
D	20	10	400	100	200
E	25	14	625	196	350
E	73	43	1339	425	737

$$\text{AŒX } Y - (\text{EX})(\text{E } Y)$$
$$\text{^}NEX^2 - (EX)^2 \text{ ^}NEY^2 - (\text{E } Y)^2$$

$$\frac{5(737) - (73)(43)}{\text{^}5(133\,9) - (73)^2 \text{ ^}5(425) - (43)^2} = \frac{3685 - 3139}{71366\ 7276}$$

$$\frac{546}{(37)(16.6)} \quad \frac{546}{614} = .89$$

Cum se calculează coeficienții de regresie:

$$EX\text{^}\_ \frac{(EX)(\text{E } Y)}{N} b = \frac{}{EX^2 - (\text{E } X)} \quad a = E \frac{Y}{N} - b \frac{\text{E}X}{N} 4$$

$$b = \frac{73 - \frac{(73)(43)}{5}}{1339 - \frac{(73)^2}{5}} = \frac{737 - 628}{1339 - 1066} = \frac{109}{273} = .40$$

$$a = \frac{43}{5} - .40\frac{(73)}{5} = 8.60 - 5.84 = 2.76$$

Anexa E
AJUSTARE-VÂRSTA

Luați în considerare două populații, A și B, cu următoarele caracteristici:

Vârsta Populației		Vârstă - specifice Rate	# de în numărul de decese în deces Populație Populație	Rata	brut
O	**Tineri**	$\frac{4}{1000} = .004$	500	.004 x 500 = 2	
	Vechi	$\frac{16}{1000} = .016$	500	.016 x 500 = 8	
	Total		1000	10	10 1000
B	**Tineri**	$\frac{5}{1000} = .005$	667	.005 x 667 = 3.335	
	Vechi	$\frac{20}{1000} = .020$	333	.020 x 333 = 6.665	
	Total		1000	10	10 1000

Rețineți că populația B are rate de deces specifice vârstei mai mari în fiecare grup de vârstă decât populația A, dar ambele populații au aceeași rată brută de deces de 10/1000. Motivul pentru aceasta este că populația A are o proporție mai mare de bătrâni (50%) și chiar dacă rata mortalității la bătrâni este de 16/1000 în populația A față de 20/1000 în populația B, numărul mai mare de persoane din acel grup contribuie la un număr mai mare de decese totale.

Pentru a efectua ajustarea vârstei, trebuie să selectăm o populație standard cu care vom compara atât A cât și B. Următoarele exemple folosesc două populații standard diferite ca ilustrații. În practică, o populație standard este aleasă fie ca populație într-un anumit an, fie ca populație combinată A și B. Alegerea populației standard nu contează. Expresia „populație standard" în acest context se referă la o populație cu o anumită distribuție de vârstă (dacă ne ajustăm pentru vârstă) sau distribuție pe sexe (dacă ne ajustăm pentru sex). Ratele specifice vârstei (sau specifice sexului, dacă acesta este ceea ce este ajustat) pentru ambele grupe A și B sunt aplicate distribuției de vârstă a populației standard pentru a compara A și B *ca și cum ar* avea aceeași distribuție de vârstă.

Rețineți că dacă utilizați două populații standard diferite, obțineți rate diferite ajustate în funcție de vârstă, dar cifrele relative sunt aceleași, adică ratele ajustate în funcție de vârstă pentru A sunt mai mici decât pentru B. Aceasta înseamnă că ratele specifice vârstei pentru A sunt mai mici decât pentru B, dar deoarece ratele brute sunt aceleași, trebuie să însemne că populația A este mai în vârstă. Deoarece știm că ratele specifice vârstei pentru persoanele în vârstă sunt mai mari decât pentru persoanele mai tinere, populația A trebuie să fi fost ponderată de o proporție mai mare de persoane în vârstă (care au contribuit la mai multe decese) pentru a avea ca rezultat aceeași rată brută ca B, dar la o rată mai mică ajustată în funcție de vârstă.

Există excepții de la concluzia de mai sus atunci când luăm în considerare grupurile în care mortalitatea infantilă este foarte mare. În acest caz, s-ar putea ca tinerii să aibă rate de deces foarte mari, chiar mai mari decât bătrânii. În societățile industrializate, totuși, ratele de deces specifice vârstei pentru bătrâni sunt mai mari decât pentru tineri.

POPULAȚIA STANDARD I: Exemplu (mai mulți bătrâni decât tineri)

Vârstă	# de Oameni	Aplicați ratele de deces specifice vârstei pentru populația A la populația standard	Numărul de decese așteptate în A dacă ar fi aceeași compoziție de vârstă ca și Populația standard	Aplicați ratele de deces specifice vârstei pentru Populația B până la Populația standard	Numărul de decese așteptate în B dacă ar fi aceeași compoziție de vârstă ca și Populația standard
Tineri	300	x .004 =	1.2	.005	1.5
Vechi	700	x .016 =	11.2	.020	14.0
Total	1000		12.4		15.5
Tarife ajustate în funcție de vârstă pentru: A B = 15/1000 = 12/1000					

POPULAȚIA STANDARD II: Exemplu (mai mulți tineri decât bătrâni)

Vârstă	# de persoane	Aplicați ratele de deces specifice vârstei pentru populația A la populația standard	Număr de decese așteptate în A dacă ar fi aceeași compoziție de vârstă ca și populația standard	Aplicați ratele de deces specifice vârstei pentru populația B la populația standard	Numărul de decese așteptate în B dacă ar fi aceeași compoziție de vârstă ca și Stan dard Popu lation
Tineri	1167	x .004 =	4,67	.005	5,84
Vechi	833	x .016 =	13.33	.020	16.66
Total	2000		18		22.50
Tarife ajustate în funcție de vârstă pentru: $\frac{18}{2000} \quad \frac{9}{1000}$			Tarife ajustate în funcție de vârstă pentru: $\frac{22.5}{2000} \quad \frac{11.25}{1000}$		

Anexa F

LIMITE DE ÎNCREDERE PRIVIND RAPORTELE DE COTE

Limitele de încredere de 95% pentru un raport de cote (OR) sunt

$$OR \times e^{\left[\pm 1.96\sqrt{\frac{1}{a}+b+c+d}\right]}$$

Pacienți cu cancer pulmonar		Controale corelate cu alte boli	
475	*a*	431	*b*
7	*c*	61	*d*

Reproducem aici tabelul din Secțiunea 4.12 pentru a fi folosit ca exemplu:

$$OR = \frac{475 \times 61}{431 \times 7} = 9.6$$

Fumători de 14-24 de ani
Tigari zilnic

nefumători

(persoane cu boala) (persoane fara boala)

Limita superioară de încredere de 95% =

$$SAU \times e^{(1.96 \times .405)} = 9.6 \times e^{.794} = 9.6 \times 2.21 =$$

2 1 . 2

Limită inferioară de încredere de 95% =

$$SAU\ x\ e^{(.196\ x\ .4\ 05)} = 9\ .\ 6\ x\ e^{-\ 794} = 9\ .\ 6\ x\ .\ 45 = 4\ .\ 3$$

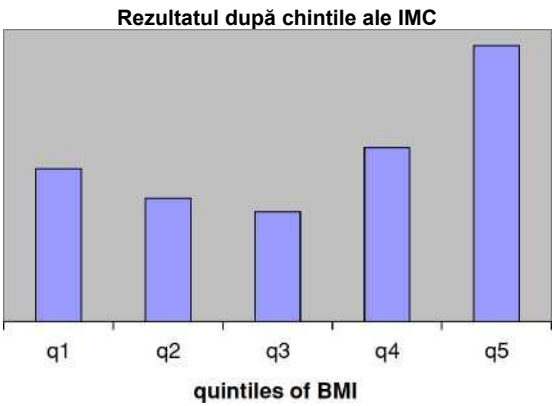
$$Notă:\ e^{-J9}{}_{4} = 44 = .\ 45$$
$$e$$

Astfel, intervalul de încredere este de la 4,3 la 21,2.

Anexa G
RELAȚIA ÎN FORMA DE „J” SAU „U” ÎNTRE DOUĂ VARIABILE

Uneori ne putem gândi că o variabilă este legată de un rezultat în formă de J sau U, de exemplu greutatea și mortalitatea. (Indicele de greutate pe care îl folosim este indicele de masă corporală - IMC - care este greutatea în kilograme împărțită la pătratul înălțimii în metri). O curbă „J” sau „U” poate fi imaginată prin reprezentarea indicelui de masă corporală în funcție de mortalitate. Forma J sau U înseamnă că pe măsură ce IMC crește, rata mortalității crește, dar rata mortalității crește și la valori foarte scăzute ale IMC, de unde forma J sau U. Acest lucru se poate datora unei boli preexistente: oamenii care sunt foarte slabi pot fi slabit deoarece sunt deja bolnavi si, desigur, vor avea mai multe sanse sa moara. Sau poate fi din cauza consecințelor fiziologice ale greutății foarte mici. Putem testa dacă există o curbă J incluzând un termen pătrat (cunoscut și ca termen patratic) în ecuația Cox.

Figura de mai jos arată un exemplu de relație în formă de J între IMC și un anumit rezultat.



După cum puteți vedea, curba are o formă de U. Atât cei cu niveluri mai mici, cât și mai mari ale IMC au niveluri mai ridicate ale rezultatului.

Mai jos sunt coeficienții [5]din SHEP (Programul de hipertensiune sistolică la vârstnici). dintr-o ecuație de regresie Cox pentru deces în rândul persoanelor din grupul de tratament activ. Un termen patratic pentru IMC a fost introdus în modelul Cox, împreună cu o serie de alte covariabile, rezultând următorii coeficienți pentru IMC:

	Coefficient	Valoarea p erorii standard
BMI	-0.3257	0,1229 ,008
BMI^2	0.0059	0,0020 , 003

Deoarece *coeficientul pentru IMC pătrat este semnificativ* , știm că relația dintre IMC și mortalitate nu este liniară.

1. **Pentru a calcula riscul relativ de deces pentru o anumită valoare a IMC în comparație cu o altă** valoare a IMC, trebuie să facem următoarele:

$$RR = e^{k} \text{ (vezi secțiunea 4.20)}$$

Pentru a compara un IMC de 32 cu un IMC de 27 dacă coeficientul pentru termenul pătrat este semnificativ,

k = coeficient de termen liniar x (IMC_1 - IMC_2) + coeficient de termen pătrat ($IMC_1{}^6$- $IMC_2{}^2$)

$k = -0{,}3257\ (32 - 27) + 0{,}005\ (32^2 - 27^2) = .1120$

$e^k = 1{,}12$

persoanele cu un IMC de 32 au cu 12% mai multe șanse de a muri decât cele cu un IMC de 27.

{Notă: dacă coeficientul termenului pătrat nu a fost semnificativ, atunci k

5Wassertheil-Smoller S, Fann C, Allman RM, Black HR, Camel GH, Davis B, Masaki K, Pressel S, Prineas RJ, Stamler J, Vogt TM, For the Grupul de Cercetare Cooperativă SHEP. Relația dintre masa corporală scăzută și deces și accident vascular cerebral în Programul de hipertensiune sistolică la vârstnici (SHEP). *Arhivele de Medicină Internă* 2000;160(4):494-500.

= coeficient liniar [IMC $_1$ - IMC $_2$]}

2. **Pentru a calcula nadirul (punctul cel mai de jos) al curbei J sau U:**

NADIR = -1/2(coeficient liniar/coeficient pătratic)
= -1/2 (-0,3257/0,0059)
= -1/2 x (-55,20) = -27,6 (rotunjit)
Aceasta înseamnă că cea mai scăzută mortalitate pentru acest grup are loc la un IMC de 27,6. Din cauza erorilor de rotunjire, valoarea nadir reală raportată în paginile de referință a fost 27,7.

3. **Acum ar putea dori să obținem riscul unui anumit IMC relativ la nadir.**
   a) RR al IMC = 25 comparativ cu nadir calculat ca -0,3257 (25 - 27,7) + .0059 (25 ^{2} - 27,7 ^{2}) = .0399 e ^{k} = 1,04
   b) RR al IMC = 30,4
      K = -.3257 (30.4 - 27.7) + .0059 (30.4 ^{2} - 27.7 ^{2}) = .0461
      e ^{k} = 1,05
   c) RR al IMC sub nadir și peste nadir este crescut

Anexa H
DETERMINAREA ADECVĂRII A SCHIMBA SCORURI

(1) Pentru a determina dacă scorurile de modificare sunt adecvate:

Luați în considerare un grup de 16 pacienți care au următoarele scoruri pe o scală care evaluează simptomele depresive; un retest este dat la scurt timp după aceea pentru a determina variabilitatea scorurilor în cadrul pacienților:

Tabelul H.1

# pacient	Scorul Scala Primului Test	Retesta ți
1	12	13
2	16	15
3	22	21
4	24	23
5	30	29
6	18	19
7	16	15
8	12	12
9	14	15
10	18	18
11	24	24
12	30	29
13	18	19
14	16	15
15	14	15
16	10	11
Medie	18.38	18.3

O analiză a varianței indică următoarele (Tabelul H.2):

Tabelul H.2

Sursa de Variație	SS	df	MS	F	*P* -valoarea
Pacienții	1014,7188	15	67,6479	156,8647	0,0000
Testare	0,0313	1	0,0313	0,0725	0,7915
Eroare	6,4688	15	0,4313		
Total	1021.2188	31			

$\overline{\sigma}^{2}_{\text{între pacienți}}$ = (67,65 - .43)/2 = 33,61

$\overline{\sigma}^{2}_{\text{+ eroare}}$ ^{33}.6 $^{1/(33}$ .61 + .43) = .987

Aceasta este mai mare de .5, astfel încât utilizarea scorurilor de modificare este adecvată. [Notă: o^{2}, sau varianța, este MS (pătratul mediu) din analiza varianței.]

În continuare, pacienții sunt împărțiți în două grupe; unui grup i se administrează o intervenție alimentară care durează 10 săptămâni, în timp ce celălalt grup servește ca grup de control. Scala se administrează din nou după 10 săptămâni la ambele grupuri, cu următoarele rezultate:

Tabelul H.3

Grupul de control				Grup de tratament			
Pacient #	Pretest	Post test	Schimbați scorul	Pacient #	Prete st	Post test	Schimbați scorul
1	12	13	1	9	14	7	-7
2	16	15	-1	10	18	10	-8
3	22	20	-2	11	24	7	-17
4	24	18	-6	12	30	5	-25
5	30	25	-5	13	18	10	-8
6	18	16	-2	14	16	8	-8
7	16	12	-4	15	14	4	-10
8	12	10	-2	16	10	5	-5
Medie	18.75	16.13	-2,63		18	7	-11
Variance	38,79	23.27	5.13		40.00	5.14	44,57
sd	6.23	4,82	2.26		6.32	2.27	6,68

(2) Pentru a calcula *coeficientul de sensibilitate* la schimbare, faceți o analiză cu măsuri repetate a variantei asupra scorurilor din *grupul de tratament* ; pentru a obține varianța erorii, calculați varianța scorurilor de modificare.

Tabelul H.4

Sursa de Variaţie	SS	df	DOMNIȘOARĂ	F	Valoarea *P*
Între testare/retestare	484	1	484,0000	21,4430	0,0004
În pacientii	316	14	22,5714		
Total	800	15			

Coeficient de sensibilitate = varianța scorurilor de modificare în grupul de tratament/(varianța scorurilor de modificare + varianța de eroare) =

$$\frac{44\ 57}{(44 . 57 + 22 . 29)} = .67$$

(44,57 este obținut din ultima coloană a tabelului G.3).

(3) *Mărimea efectului* = media scorurilor de modificare/ sd a scorurilor pretest:

în grupul de tratament = -11/6,32 = -1,74 (a existat o *scădere a scorului simptomelor depresiei* de 1,74 unități de abatere standard pretest).

(4) *Măsura de răspuns* [25 a lui Guyatt] este (scorurile medii ale modificării în grupul de tratament)/(sd ale scorurilor modificării la subiecții stabili). Presupunem aici că grupul de control este grupul subiecților stabili, deși, în general, „subiecții stabili" se referă la subiecții care sunt stabili în raport cu un anumit criteriu extern.

$$\begin{matrix} -11 \\ 2.26 \end{matrix} \qquad - 4.86$$

(5) *Comparație cu un grup de control:* Mărimea efectului pentru grupul de tratament este -1,74, deci depășește în mod clar modificarea grupului de control, care este -2,63/6,23 = -.42. Dacă calculăm raporturile dintre tratament și grupul de control pentru indicii de receptivitate de mai sus, vom constata în acest exemplu că aceștia sunt foarte similari.

Pentru dimensiunea efectului, raportul este -1,74/-.42 = 4,14. Pentru statistica lui Guyatt este -4,86/-1,16 = 4,19. (-1,16 a fost obținut prin modificarea medie în grupul de control împărțit la abaterea standard a scorurilor de schimbare în grupul de control; adică = -2,63/2,26.)

Pentru coeficientul de sensibilitate, acesta este .67/.14 = 4.78. (Valoarea 0,14 a fost obținută făcând o analiză a scorurilor din grupul de control, care nu este prezentată aici, așa că luați-l pe credință sau calculați-l ca o verificare a acurateței acesteia.)

Anexa I
PRINCIPII GENETICE

Totul ar trebui făcut cât se poate de simplu, dar nu mai simplu.
(atribuit lui Albert Einstein, 1879-1955)

1. ADN-ul (acidul dezoxiribonucleic) este format din patru unități – sau maree nucleo – care sunt molecule compuse din carbon, hidrogen, oxigen și fosfor. Aceste molecule, numite baze, sunt adenina, guanina, timina si citozina si sunt notate cu literele A, G, T si C.
2. ADN-ul este aranjat în două catene răsucite într-o formă de dublă helix, astfel încât nucleotidele AGCT se împerechează între ele în moduri fixe. Un A se perechează întotdeauna cu T și C se face întotdeauna perechi cu G. Acestea se numesc perechi de baze. Dacă un fir din elica dublă ar fi înșirat într-o linie, ar putea arăta astfel:

AATTCGTCAGTCCC, Cealaltă componentă care se asociază cu ea ar fi: TTAAGCAGTCAGGG.

3. Există 3 miliarde de astfel de perechi de baze (sau 6 miliarde de baze) în genomul uman (care se referă la tot materialul genetic la om). Aceste 3 miliarde de perechi de baze sunt organizate în 23 de perechi de cromozomi (una de la mamă și una de la tată), care se află în fiecare celulă vie din organism.
4. Din aceste 3 miliarde de perechi de baze există aproximativ 30.000 de gene (estimarea variază – în 1970 era de 300.000; în anii 1990 era de 100.000; ulterior, după decodificarea genomului uman, s-a crezut că ar fi de 30.000, estimarea a fost din nou mărită mai recent; 45.000), care sunt secvențe de perechi de baze de lungimi diferite. Secvențele rămase sunt în cea mai mare parte ADN „junk", deși unele secvențe, nu genele în sine, servesc funcții de reglare.

5. Genele instruiesc celulei cum să producă proteine (sau lanțuri polipeptidice) care sunt compuse din aminoacizi.
6. Un grup de 3 baze se numește codon. Un codon codifică un singur aminoacid. Proteinele constau din mai mulți sau mai mulți aminoacizi.

Deoarece există 4 litere diferite (nucleotide), s-ar crede că ar trebui să existe 4x4x4x4 sau 64 de aminoacizi diferiți, dar sunt doar 20 (nu întrebați de ce). Baza întregii vieți sunt acești 20 de aminoacizi.

7. Genele produc proteine prin „codificarea" aminoacizilor. Aceasta înseamnă că gena direcționează ansamblul moleculei proteice prin - specificarea secvenței de aminoacizi și a modului în care aceștia ar trebui să se lege pentru a produce proteina particulară, o enzimă, de exemplu, care transformă colesterolul în estrogen sau un neurotransmițător sau os.
8. O grămadă de codoni (fiecare alcătuit din 3 baze și fiecare codifică unul dintre cei 20 de aminoacizi) se numește exon. Astfel, un exon este compus din mulți codoni, fiecare codifică un singur aminoacid și împreună codifică o proteină specifică.
9. Între exoni dintr-o genă există și alte secvențe de litere numite introni. Intronii sunt secțiuni de ADN (întinderi ale literelor A,G,C,T) care nu codifică aminoacizi. Nu se știe care sunt funcțiile intronilor, dacă există.
10. O genă este formată din exoni și introni. Înainte de a putea direcționa efectiv fabricarea unei proteine, aceasta este transcrisă în nucleul - celulei în ARN (acid ribonucleic), care este o copie exactă a unei singure catene a ADN-ului, (cu excepția faptului că în loc de baza timină, T, conține uracil, U). Acest ARN iese apoi din nucleu și în corpul (citoplasma) celulei unde vor fi fabricate proteinele. Lucrul incredibil este că apoi părțile ARN-ului care nu sunt exoni (sau regiuni codificatoare) sunt îndepărtate și acesta se numește acum ARNm (ARN mesager). Așadar, la sfârșitul acestui proces de transcripție, toți intronii sunt despărțiți și rămân doar exonii (regiunile codificatoare), gata să direcționeze materialele în interiorul celulei, în afara nucleului, pentru a produce lanțul de aminoacizi care constituie proteina specifică codificată de acea genă particulară. Odată ce proteina este fabricată, este fie folosită de celulă, fie transportată în vezicule către membrana celulei, unde este expulzată în țesut sau sânge.
11. Când spunem că o genă este „exprimată", înseamnă că se formează ARNm și o proteină. Doar aproximativ 3% din genom (adică 3% din tot ADN-ul) este exprimat; restul este fie de reglementare, fie este junk (din câte știm acum).
12. Deoarece nu tot ADN-ul este compus din gene, găsirea de gene

încorporate în tot acel ADN este o provocare majoră. Dar haideți să spunem că găsiți o secvență de baze care este ATG, care codifică aminoacidul metionină. Acesta este, de asemenea, cunoscut sub numele de codonul de pornire, deoarece îi spune celulei să înceapă să-și producă proteina din acel moment; celula continuă să producă aminoacizii în secvența specificată de codoni, până când întâlnește codonul Stop care are una dintre cele trei forme: TAA, TAG sau TGA, moment în care se oprește.

13. Există multe redundanțe în sistem, iar cea de-a treia poziție este „clintită", ceea ce înseamnă, de exemplu, că, chiar dacă în loc de TCA, ai un TCC, ai putea totuși să faci același aminoacid (serină). Pe de altă parte, s-ar putea ca o singură schimbare a unei litere, să zicem că ai un T unde majoritatea oamenilor au un C, ar putea distruge întreaga proteină deoarece lipsește un aminoacid sau se face altul.
14. Aceste variații între oameni dintr-o singură literă a codului sunt cunoscute ca polimorfisme cu un singur nucleotide (SNP, pronunțate ca snips). Se estimează că poate exista aproximativ un SNP la 1000 de baze de ADN. Să presupunem că SNP poate avea două forme – adică două alele, adică un anumit punct ar putea avea, de exemplu, nucleotida A sau T. Dacă te uiți la o întindere de ADN cu o lungime de 2000 de baze, ar putea exista 4 variații de SNP, 2 în primele 1000 de baze și 2 în a doua 1000 de baze. Unele SNP sunt variante normale într-o populație, unele pot predispune la boală, iar altele pot provoca boli.
15. Polimorfismele sunt, așadar, variante ale unei gene sau variante într-un anumit loc sau loc în interiorul unei gene, dintre care majoritatea sunt variante inofensive . Un polimorfism particular, de exemplu, poate consta din două forme alternative (alele) la un loc particular; de exemplu, puteți avea fie un A, fie un C la acel loc. Polimorfismele pot consta, de asemenea, în inserții ale unei nucleotide într-un anumit loc, sau o ștergere a unei nucleotide sau alte variații.
16. Oamenii de știință încep să se uite la blocuri de ADN - porțiuni lungi de ADN într-o anumită locație a unui cromozom care au un model distinctiv de SNP. Acestea se numesc haplotipuri. Haplotipurile de interes sunt secvențe de alele care sunt apropiate pe același cromozom și, prin urmare, tind să fie moștenite ca bloc. Este posibil să se analizeze asocierea cu boala SNP-urilor care diferențiază unele modele comune de haplotipuri, mai degrabă decât să se analizeze asocierea fiecărui

SNP individual cu boala. Construirea hărților haplotipurilor („hapmaps") este un domeniu activ de cercetare care arată mult promițător în a putea identifica polimorfismele bolii mai eficient.

17. Întregul genomul uman constă apoi din aproximativ 30.000 de gene plus tot restul materialului genetic care nu au nicio funcție sau nu au nicio funcție , care alcătuiesc cele 3 miliarde de perechi de baze. Fiecare celulă nucleată din organism are aceleași 3 miliarde de perechi de baze (cu excepția spermatozoizilor și a ovulului, fiecare dintre ele având jumătate din aceasta). Deci întrebarea ar putea fi dacă fiecare celulă are același ADN, de ce nu fiecare celulă produce aceleași proteine? De ce o celulă hepatică nu produce insulină? De ce o celulă a creierului nu face os? Acest lucru se datorează faptului că o parte a unei gene, sau uneori o secvență aflată la o anumită distanță de genă, îi spune genei când să pornească sau să oprească, adică reglează expresia genei.
18. În timpul diviziunii celulare (mitozei), în toate celulele, cu excepția spermatozoizilor și a oului, are loc mai întâi o duplicare a cromozomilor și apoi o segregare a părților duplicate, astfel încât atunci când celula se împarte în jumătate, cele două noi celule fiice vor avea fiecare un set complet de 23 de perechi de cromozomi. Cei doi membri ai fiecărei perechi poartă copii ale acelorași gene (și sunt omologi), dar pot diferi în anumite puncte (loci) de pe cromozom, deoarece un membru al perechii provine de la mamă și unul de la tată.
19. În cazul celulelor germinale din testicule sau ovare, care sunt destinate să devină spermatozoizi și ovul, diviziunea celulară se numește meioză, iar în acest tip de diviziune celulară nu există o duplicare a cromozomilor înainte de diviziune, astfel încât atunci când spermatozoizii și ovulul se întâlnesc, rezultă doar un set de cromozomi (un cromozom din fiecare pereche din ovul și provine unul din spermă). În timpul meiozei, cei doi cromozomi omologi din perechea de cromozomi sunt foarte aproape unul de celălalt și bucăți mici de ADN trec de la un cromozom al perechii la celălalt cromozom al perechii. Acum amintiți-vă că aceasta se află în celula germinativă (celula precursoare a spermei și a oului). Această celulă germinală se împarte (prin meioză, mai degrabă decât prin mitoză) Deci, când spermatozoidul este în sfârșit format, are doar un membru al fiecărei perechi de cromozomi, dar acest singur cromozom al perechii este

puțin diferit de cromozomul original care se află în fiecare celulă a corpului părinte, din cauza schimbului de ADN care a avut loc în timpul meiozei. (Acest lucru se numește recombinație). Ceva similar se întâmplă cu ovulul, iar când cei doi se întâlnesc în fecundare și formează un gamet (o singură celulă fertilizată din care se va dezvolta organismul uman) această celulă are acum un set complet de 23 de perechi de cromozomi, câte una în fiecare pereche de la mamă și una de la tată.

20. Recombinarea este importantă în descoperirea genelor legate de boală din cauza următoarelor: dacă alelele de la doi loci de pe cromozom sunt apropiate una de cealaltă, ele vor fi moștenite împreună, dar dacă sunt suficient de îndepărtate, este mai probabil să se recombine (încrucișate) în timpul meiozei și vor fi moștenite independent una de cealaltă. Această stare de lucruri ajută oamenii de știință să localizeze genele. Recombinarea datorită încrucișării unei părți a ADN-ului de la un cromozom al unei perechi la celălalt, are loc numai în timpul meiozei în celula germinativă. După aceea, de fiecare dată când o celulă se reproduce, și o fac tot timpul, se împarte în două cu exact aceleași copii ale complementului de 23 de perechi de cromozomi.
21. Mutațiile sunt erori în replicarea ADN-ului. Ele apar spontan, foarte rar - aproximativ 1 din 1.000.000.000 de baze, dar expunerile mediului pot provoca mutații. Mutațiile pot fi somatice, adică în celulele organismelor vii, caz în care se sting atunci când organismul moare, sau în celulele germinale care devin spermatozoizi sau ovule și în care caz se transmit de-a lungul generațiilor. Mutațiile pot fi de mai multe tipuri și pot să nu aibă niciun efect, de exemplu dacă au loc într-un intron, sau pot schimba aminoacidul produs și astfel pot altera proteina . O mutație poate șterge o singură nucleotidă, introduce o singură - nucleotidă sau înlocuiește o nucleotidă cu alta, să spunem înlocuirea unui C cu un T. O mutație de inserție sau deleție poate provoca o „schimbare a cadrului", ceea ce înseamnă că începutul și/sau sfârșitul regiunii de codificare ar fi diferit și astfel s-ar forma o secvență diferită de aminoacizi. Într-o „mutație missens" unul dintre cei 20 de aminoacizi este înlocuit cu altul (modificând astfel proteina); într-o mutație „prostii", apare prematur un codon stop. După cum s-a menționat anterior, unele mutații nu au niciun efect asupra organismului și unele pot cauza sau predispune la boli.

De fapt, este miraculos că atunci când atât de multe pot merge prost, atât de puțin se întâmplă de fapt!

REFERINȚE

Capitolul 1

1. Popper KR: Logica descoperirii științifice. New York: Harper and Row, 1959.
2. Weed DL: Despre logica inferenței cauzale. Jurnalul American de Epidemiologie, 123(6):965-979, 1985.
3. Goodman SN, Royall R: Dovezi și cercetare științifică. AJPH, 78(12):1568-1574, 1988.
4. Susser M: Reguli de inferență în epidemiologie. În: Regulatory Toxicology and Pharmacology, Chapter 6, pp. 116-128. New York: Academic Press, 1986.
5. Brown HI: Percepție, teorie și angajament: noua filozofie a științei. Chicago: Precedent, 1977.
6. Hill B: Principles of Medical Statistics, ed. a 9-a. New York: Oxford University Press, 1971.

Capitolul 2

7. Arias E, Smith BL. Decese: Date preliminare pentru 2001. Rapoarte naționale de statistică vitală; vol. 51, nr. 5. Hyattsville, Maryland: Centrul Național pentru Statistică în Sănătate. 2003.

Capitolul 3

8. Drapkin A, Mersky C: Terapia anticoagulantă după infarct miocardic acut. JAMA, 222:541-548, 1972.
9. Dezvoltarea intelectuală a copiilor: Departamentul de Sănătate, Educație și Bunăstare al SUA, Serviciul de Sănătate Publică, HSMHA, decembrie 1971, Seria 11 de Statistici Vitale și de Sănătate, 10 noiembrie.
10. Davis BR, Blaufox MD, Hawkins CM, Langford HG, Oberman A, Swencionis C, Wassertheil-Smoller S, Wylie-Rosett J, Zimbal-di N:

Trial of Antihypertensive Interventions and Management. Design, Metode și Rezultate de referință selectate. Controlled Clinical Trials, 10:11-30, 1989.
11. Oberman A, Wassertheil-Smoller S, Langford H, Blaufox MD, Davis BR, Blaszkowski T, Zimbaldi N, Hawkins CM, pentru grupul TAIM: Tratamentul farmacologic și nutrițional al hipertensiunii ușoare: modificări ale stării de risc cardiovascular. Analiști de Medicină Internă, ianuarie 1990.
12. Rothman KJ: Nu sunt necesare ajustări pentru mai multe comparații . Epidemiology, 1(1):43-46, 1990.
13. Scarr-Salapateck S: Race, Social Class, and IQ Science, 174: 1285-1295, 1971.
14. Sokal RR, Rohlf JF: Biometrie. San Francisco: WH Freeman, 1 969.

Capitolul 4

15. NCHS: Raport avansat al statisticilor finale ale mortalității, 1987, Supl. (pag. 12), MSVR Vol. 38, nr 5, Pub. Nr. (PHS) 89-1120, Hyattsville, MD: Public Health Service, 26 septembrie 1989.
16. Grupul de cooperare al programului de detectare și urmărire a hipertensiunii arteriale: studii asupra tensiunii arteriale în 14 comunități. Un ecran în două etape pentru hipertensiune arterială. JAMA, 237(22):2385-2391, 1977.
17. Comisia Inter-Societate pentru Resurse pentru Bolile Inimii. Grupul de studiu pentru scleroza aterosclerozăși Grupul de studiu pentru epidemiologie: Prevenirea primară a bolilor aterosclerotice. Circulație, 42:A55, 1 970.
18. Kannel WB, Gordon T, eds. Studiul Framingham: o investigație epidemiologică a bolilor cardiovasculare. februarie 1974, DHEW Publ. Nr. (NIH) 74-599.
19. Aceste date provin din Proiectul Național de Pooling. În sensul acestui exemplu, hipertensiunea arterială este definită ca tensiune arterială diastolică > 105 mm Hg, iar „normal” este definit ca PB < 78 mm Hg. Boala în cauză este un „Eveniment coronarian” și perioadele de

timp sunt de 10 ani. Rețineți că hipertensiunea arterială este definită în prezent ca tensiune arterială sistolică > 140 mmHg și/sau tensiune arterială diastolică > 90 mmHg. Notă: limitele actuale pentru - hipertensiune arterială sunt o tensiune arterială sistolică > 140 mmHg sau o tensiune arterială dialostică > 90 mmHg.

20. Lilienfeld AM: Fundamentele epidemiologice, p. 180. New York: Oxford University Press, 1976.
21. Grup de redactare pentru anchetatorii inițiativei pentru sănătatea femeilor. Riscuri și beneficii ale estrogenului plus progestativ la - femeile sănătoase după menopauză. JAMA-Express, 288:321-333, 2002.
22. D'Agostino Jr. RB. Tutorial in Biostatistics Propensity Score Methods for Bias Reduction in the comparison of a treatment to a non-randomized control group. Statistics in Medicine, 17:2265-2281, 1998.
23. Exemplul numeric este prin amabilitatea Dr. Martin Lesser, Cornell University Medical Center.
24. Greenland S. Modeling and Variable Selection in Epidemiologic Analysis. AJPH, 79(3):340-349.

Capitolul 5

25. Meyer KB, Pauker SG: Screening pentru HIV: Ne putem permite rata fals pozitive? New England Journal of Medicine, 317(4): - 238-241, 1987.

Capitolul 6

26. Grupul de cercetare al Proiectului de medicamente coronariene: Influența aderenței la tratament și răspunsul colesterolului asupra mortalității în proiectul de medicamente coronariene. New England Journal of Medicine, 303:1038-1041, 1980.
27. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG: Proiectarea și analiza studiilor clinice randomizate care necesită observarea prelungită a fiecărui pacient. I. Introducere și Proiectare. British Journal of Cancer, 34:585-612, 1976.

Capitolul 7

28. Guyatt G, Walter S, Norman G: Măsurarea schimbării în timp: evaluarea utilității instrumentelor de evaluare. Journal of Chronic Diseases, 40:171-178, 1987.

Capitolul 8

29. Ellsworth DL, Manolio TA. Importanța emergentă a geneticii în cercetarea epidemiologică. I. Concepte de bază în genetică umană și tehnologie de laborator. Ann Epidemiol., 9:1-16, 1999.
30. Ellsworth DL, Manolio TA. Importanța emergentă a geneticii în cercetarea epidemiologică II. Probleme în proiectarea studiului și cartografierea genelor. Ann Epidemiol. 9:75-90, 1999.
31. Ellsworth DL, Manolio TA. Importanța emergentă a geneticii în cercetarea epidemiologică. III. Bioinformatică și metode genetice statistice. Ann Epidemiol., 9:207-224, 1999.
32. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M și colab. Factori de mediu și ereditari în cauzarea cancerului - analize ale cohortelor de gemeni din Suedia, Danemarca și Finlanda. N Engl J Med 2000;343(2):78-85.
33. Plomin R, Owen MJ, McGuffin P. Baza genetică a comportamentelor umane complexe. Science, 264:1733-1739, 1994.
34. Risch N, Merikangas K. Viitorul studiilor genetice ale bolilor umane complexe. Science 1996;273(5281):1516-7.
35. Owen MJ, Cardno AG și O'Donovan MC. Genetica psihiatrică : înapoi în viitor. Molecular Psychiatry 2000;5:22-31.
36. Risch NJ. În căutarea determinanților genetici în noul mileniu . Nature, 405:847-856, 2000.
37. Burmeister M. Genetica complexă și implicații pentru psihie încearcă. Concepte de bază în studiul bolilor cu genetică complexă . Biol Psychiatry, 45:522-532, 1999.
38. Schork NJ. Genetica bolilor complexe. Abordări, probleme și soluții. Am J Respir Crit Care Med. 156:S103-S109, 1997.

LECTURI SUGERATE

Anderson S, Auquier A, Hauck WW, Oakes D, Vandaele W, Weisberg HI: Statistical Methods for Comparative Studies. New York: John Wiley and Sons, 1980.

Nivel intermediar. Necesită un curs anterior de statistică. Excelent pentru cercetătorii aplicați.

Balding DJ, Bishop M, Cannings C, editori. Manual de Genetică Statistică. West Sussex, Anglia, John Wiley & Sons, editori, 2001.

O carte de referință complexă.

Campbell DT, Cook TD: Quasiexperimentation: Design and Analysis for Field Settings. Chicago: Rand McNally, 1979.

Se concentrează pe problemele de proiectare și măsurare, în special pe tipurile de părtinire, care pot apărea în studiile cvasi-experimentale. De asemenea, se ocupă de tehnicile statistice adecvate care trebuie utilizate. Deosebit de relevant pentru persoanele interesate de evaluarea programului.

Cohen J: Analiza puterii statistice pentru științe comportamentale, ed. a 2-a. Hillsdale, NJ: Lawrence Erlbaum, 1988.

Un clasic și o necesitate. Tot ce ați dorit vreodată să știți despre calculele dimensiunii eșantionului, explicat clar și cuprinzător. O carte de referință și sursă. Probabil trebuie interpretat de un statistician.

Cohen P (ed.), Cohen J, West SG, Aiken LS: Applied Multiple Regres sion: Corelation Analysis for the Behavioral Science, ed. a 3-a, Law rence Erlbaum, 2002

Expunere excelentă, amănunțită și clară a subiectelor de top foarte complexe. Este destul de avansat.

Elwood JM: Relații cauzale în medicină. Un sistem practic de evaluare critică. New York: Oxford University Press, 1988.

Carte excelentă, cu explicații clare despre modelele de studiu, - conceptele epidemiologice și metodele statistice relevante.

Fleiss JL: Statistical Methods for Rates and Proportions, ed. a 2-a. New York: John Wiley and Sons, 1982.

Un excelent text clasic de statistică de nivel doi, care se referă la analiza datelor calitative sau categoriale.

Fleiss JL: Proiectarea și analiza experimentelor clinice. New York: John Wiley and Sons, 1986.

Se concentrează pe aspectele tehnice ale proiectării și analizei statistice a studiilor experimentale.

Fryer HPC: Concepte și metode de statistică experimentală. Boston: Allyn & Bacon, 1966.

Carte de bază detaliată despre metode statistice. Nivel intermediar spre superior. Multe formule și simboluri. Statistici mai aprofundate.

Haines JL și Pericak-Vance MA: Abordări ale cartografierii genelor în bolile umane complexe. New York: Wiley-Liss, 1998.

Hill, B: Metode statistice în medicina clinică și preventivă. Edinburgh și Londra: ES Livingstone, 1962.

Text clasic pe subiect.

Hosmer DW, Lemeshow S: Regresia logistică aplicată. New York: John Wiley and Sons, 1989.

Carte avansată despre construirea de modele în regresia logistică, necesită cunoștințe statistice.

Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH: Biostatistics in Clinical Medicine. New York: Macmillan, 1983.

Adresat mai ales medicilor. Nivel intermediar. Foarte bune explicații ale biostatisticii. Multe exemple din literatura de specialitate.

Kahn HA: O introducere în metodele epidemiologice. New York: Oxford University Press, 1983.

Carte de nivel intermediar despre metode epidemiologice și statistice selectate. Explicații bune despre analiza tabelului de viață și regresie multiplă și funcții logistice multiple. Explicații clare ale studiilor longitudinale folosind anii persoană.

Kelsey JL, Thompson WD, Evans AS: Metode în Epidemiologie observațională. New York: Oxford University Press, 1986.

Explicații extrem de utile ale problemelor implicate în studiile caz-control, retrospective și prospective. O discuție bună despre potrivire , stratificare și probleme de proiectare.

Kleinbaum DG, Kupper LL: Analiza de regresie aplicată și alte metode cu mai multe variabile. Belmont, CA: Wadsworth, 1978.

Acest text, după cum indică titlul, tratează regresia multiplă și subiecte conexe.

Kleinbaum DG, Kupper LL, Morgenstern E: Principii de cercetare epidemiologică și metode cantitative. Belmont, CA: Wadsworth, 1982.

Un text avansat conceput în primul rând pentru persoanele care efectuează cercetări epidemiologice observaționale. Sunt acoperite atât problemele de proiectare, cât și cele statistice.

Meinert CL în colaborare cu Tonascia S. Clinical Trials: Design, Conduct, and Analysis. New York, Oxford: Oxford University Press, 1986.

Ott J. Analiza legăturii genetice umane. Johns Hopkins University Press, 1999.

Popcock SJ: Studii clinice: O abordare practică. New York: John Wiley and Sons, 1983.

O carte excelentă pentru oricine care face un studiu clinic.

Riegelman RK: Studierea unui studiu și testarea unui test: Cum să citiți literatura medicală. Boston: Little, Brown, 1981.

O introducere în metodele și principiile epidemiologice destinate clinicienilor.

Rothman KJ și Greenland S: Epidemiologie modernă. a 2 -a Ed. Phila - delphia: Lippincott Williams & Wilkins, 1998.

Un text avansat care acoperă atât problemele de design, cât și cele statistice. Accentul este un studiu epidemiologic observațional și este îndreptat mai mult către cercetător decât către clinician.

Sackett DL, Haynes RB, Tugwell P: Epidemiologie clinică: O știință de bază pentru medicina clinică. Boston: Little, Brown, 1985.

Accentul este pus pe aplicațiile epidemiologiei în practica clinică.

Sham P: Statistici în genetica umană. Londra, 1998

Siegel S și Castellan Jr. NJ: Statistici neparametrice pentru științe comportamentale. Ediția a 2 -a · New York, Toronto, Londra: McGraw Hill, Inc., 1988.

O carte „cum se face”; o referință excelentă; conturează fiecare procedură pas cu pas, un clasic.

Sokal RR, Rohlf JF: Biometrie, ed. a 3-a. San Francisco: WH Freeman, 1995.

Destul de complex și cuprinzător.

Streiner L, Norman GR: Scale de măsurare a sănătății. Un ghid practic pentru dezvoltarea și utilizarea lor. New York: Oxford University Press, 1989.

Carte excelentă și de înțeles despre construcția scalei și schimbarea scorurilor.

Szklo M, Nieto FJ: Epidemiologie. Dincolo de elementele de bază. Maryland: As pen Publishers, 2000.

Undeva între textul de nivel întâi și al doilea, care se distinge prin stilul său clar și foarte lizibil, și multe exemple excelente.

INDEX

Risc absolut, 103
Precizia testului, 131
Legea aditivă a probabilității, 27
Efecte de interacțiune a modelului aditiv, 124-128
Rate ajustate în funcție de vârstă, 90-92 Ajustare în funcție de vârstă, 207-209 Rata de mortalitate anuală specifică vârstei, 90
Ipoteza alternativă, 11
Alele, 173
Abordarea partajării alelelor, 179
Eroare alfa, 16
Ipoteza alternativă, 145-146
Aminoacizi, 222
Analiza variației (ANOVA), 69-72
  cu doi factori, *vezi* ANOVA cu doi factori
ANOVA, *vezi* Analiza variației AR (risc atribuibil), 107-109 Afirmații, 5
Asociația, 78
Studii de asociere, 175-178, 180-181
Risc atribuibil (AR), 107-109
Politici de autor, 195-197

Perechi de baze, 221
Probabilitate bayesiană, 24
Curbă în formă de clopot, 40-41
Beneficență, 191
Eroare beta, 16
Varianta între grupuri, 70 procedura Bonferroni, 72-74
Gene candidate, 177, 185 Studiu caz-control, 9, 97
  calculul riscului relativ din, 104-107 Variabile categoriale, 10 Căi cauzale, 79-81 Mortalitate anuală specifică cauzei rata, 89
Diviziune celulară, 224 CentiMorgan (cM), 176
Tendință centrală, 37 Scoruri de schimbare, 161, 165
  determinarea oportunității, 217-220
Chi-pătrate, 31-32
  valori critice ale, 200 Test Chi-pătrat, 29-33 Localizare cromozomală, 176 Semnificație clinică, 153 Studii clinice, 9, 10, 141-160
  ca „etalon de aur”, 146 147
  randomizat, 141-143
  dimensiunea de, 151-153
cM (centiMorgan), 176
Codoni, 222
Coeficient de sensibilitate, 165, 219-220
Cohorta, 94, 101
Studiu de cohortă, 9, 10
Comparații între două grupuri, 60
Rate de concordanță, 174
Probabilitate condiționată, 23-24
Intervale de încredere, 53-56
Intervale de încredere *(continuare)* în jurul diferenței dintre două medii, 58-59 pentru proporții, 57-58
Limite de încredere, 55 pe raportul de cote, 211
Confuzie prin indicație, 116-117
Variabile de confuzie, 111-112
Conjecturi și respingeri, metoda, 5
Consimțământ, informat, 192-194
Tabel de urgență, 29-30

Variabile continue, 10
Grupul de control, 143, 166-167
Coeficient de corelație, calcul 78-79, conexiune 205-206 între regresia liniară și, 84 Modelul hazardelor proporționale Cox, 120-122
Valoare critică
de chi-pătrat, 32-33
din Z, 62
alfa lui Cronbach, 163
Studiu transversal, 8, 9, 97 studiu longitudinal versus, 98 101
Rata brută anuală a mortalității, 89
Punct de tăiere, 136-139

Monitorizarea datelor și a siguranței Board (DSMB), 196
Rata mortalității, 89
Inferența deductivă, 1-2
Grade de libertate (df), 56-57 Acid dezoxiribonucleic (ADN), 221
Variabile dependente, 96
Epidemiologie descriptivă, 87 df (grade de libertate), 56-57 Diferența dintre două intervale de încredere medii în jur, 58-59 calculul dimensiunii eșantionului pentru testare, 159-160 eroare standard de, 46-47 *t* -test pentru, 62-64
Diferența dintre două proporții, calculul mărimii eșantionului pentru, 158-159
Diferențele, distribuția, 46 Variabile discrete, 10 Distribuția diferențelor, 46 ADN (acid dezoxiribonucleic), 221
Studiu dublu-orb, 142 DSMB (Date și siguranță Consiliul de monitorizare), 196

Mărimea efectului, 154, 166
Epidemiologie definită, 87 descriptiv, 87 genetic, vezi Impactul epidemiologiei genetice asupra vieților umane, 197-198
molecular, 172 utilizări ale, 87-88
Epistasis, 187 Echilibrul, 194 Patrat mediu al erorii, 72
Erori, tipuri de, 14 Etică, cercetare, statistică și, 189-196
Etnie, 181
Exoni, 222
Frecvența așteptată, 20
Studii experimentale, 8

Notație factorială, 202
Rata fals-negative, 130
Rata fals pozitive, 130
Falsificare, 5
Testul exact al lui Fisher, 33, 201-202
Raportul F, 70
Frecvență, așteptată, 20

Gene, 221-222
Epidemiologie genetică, 171-187
prezentare generală a, 171-173
Analiza legăturii genetice, 175-178
Principii genetice, 221-226
Genomul, 221
Genomica, 172
Măsura de răspuns a lui Guyatt, 220

Hărți haplotip, 185-186 Haplotipuri, 224
Raportul de pericol, 121
Indicele de ereditabilitate, 174
Eterogenitate, 186

Vieți umane, impactul epidemiologiei asupra, 197-198
Subiecte de cercetare umane, protecția, 190-192
Metoda ipotetico-deductivă, 2

Rata de incidență, 93
Variabile independente, 96
Inferență inductivă, 2, 7-8
Inferență, 1
deductiv, 1-2
inductiv, 2, 7-8
Consimțământ informat, 192-194
Integritate, cercetare, 194-195
Analiza intenției de tratare, 150 151
Interacțiune(e), 124-128
între două variabile, 75-76
Introni, 222

Probabilitate comună, 21
Justiție, 192

curbele de supraviețuire Kaplan-Meier, 117
Kappa, 35-36 test Kruskal-Wallis, 77-78, 203 204

Potrivire la cele mai mici pătrate, 82
metode de tabel de viață, 117-120
Raportul de probabilitate (LR), 26
dintr-un test negativ, 131 dintr-un test pozitiv, 131
Regresia liniară
legătura între și coeficientul de corelație, 84 multiplu, 84-85
Dezechilibru de legătură, 176, 182
Statistică de legătură, 178-179 Studii de legătură, 173, 175-178 Scorul LOD, 178-179
Regresie logistică, multiplă, 113-116
Studiu longitudinal, 9, 10 studiu transversal versus, 97-101
LR, vezi Raportul de probabilitate

Pereche asortată / -test, 66-67
Potrivire, 112-113
Testul McNemar, 34-35, 184
Mijloace(e), 37
diferență între două, vezi Diferența între două mijloace
regresia către eroarea standard 147-149 a 43, 50
Pătrat mediu, 72
Median, 37
Meioza, 224
Mitoză, 224
Mod, 37
Epidemiologie moleculară, 172
Morbiditate, 93
Rata mortalității, 88-90
Regresie liniară multiplă, 84-85
Regresie logistică multiplă, 113-116
Legea multiplicativă a probabilității, 27
Efecte de interacțiune a modelului multiplicativ, 124-128
Modele multivariate, selectând variabile pentru, 122-124
Mutații, 225-226
Evenimente care se exclud reciproc, 20-21

Test negativ, 129 raportul de probabilitate al lui a, 131 valoarea predictivă a lui a, 130
Metode de legătură neparametrică,

179
Test neparametric, 67
Distribuție normală, 40-41
standardizat, 47-51
Ipoteza nulă, 6, 11-12 testare, 12-14
Studii observaționale, 8 Observații, 1
Raportul de cote (OR), 25, 106
limitele de încredere pentru, 211
Test cu o singură coadă, 64
test cu două cozi versus, 145-146
Potrivire unu-la-unu, 112 SAU, vezi Raportul cotelor

PAR (risc atribuibil populației), 107-109
Analiza legăturii parametrice, 179
Percepția pacientului, 161 Penetranță, 186
Persoană-ani de observație, 94-95
Fenocopii, 186-187 Fenotip, 186
Placebo, 141
Prevalența punctuală, 92
Polimorfisme, 223-224
Estimare cumulata, 46
Riscul atribuibil populației (PAR), 107-109
Stratificarea populației, 181 Valori ale populației, 52-53 Test pozitiv, 129
Raportul de probabilitate al lui a, 131 valoarea predictivă a lui a, 130, 132 Probabilitatea posttest de boală, 132
Putere, statistică, 154
Valoare predictivă
a unui test negativ, 130
a unui test pozitiv, 130, 132
Probabilitate pretest de boală, 132
Rata de prevalență, 92
Model probabilistic, 6-7
Probabilitate(i), 19-20
Bayesian, 24
combinând, 20-23
condiționat, 23-24
de boală, pretest și posttest, 132
articulație, 21
Analiza tendinței, 116-117
Proporții
intervale de încredere pentru, 57-58 diferență între două, calculul dimensiunii eșantionului pentru, 158-159
erori standard ale, 57
Testul Z pentru compararea a două, 60 62
Studii prospective, 9, 10, 97, 105
calculul riscului relativ din, 103-104 Proteomics, 172
Psihometrie, 164
valoarea *p , 152*

Calitatea vieții, 161-169
nevoie de standarde de cercetare în, 168-169
capcane în evaluare, construcția la scară 167-169 pentru evaluare, 162-167

Randomizare, scopurile, 143 144
Alocare aleatorie, performanță, 144-145
Studii clinice randomizate, 141-143
Interval de date, 41
Rata, 89
Caracteristica de funcționare a receptorului
curba (ROC), 139

Recombinare, 225
Referințe, 227-231
Regresia, 82-84
spre medie, 147-149
Coeficienți de regresie, 84 calcul, 205-206
linii de regresie, 82
Risc relativ (RR)
calculul, din studii prospective, 103-104
estimare a, din studii caz-control, 104-107
măsuri de, 101-103
Fiabilitate, 162-164
Repetabilitate, 162
Cercetare, nevoie de standarde de, în evaluarea calității vieții, 168-169
Etica cercetării, statistică și, 189-196
Integritatea cercetării, 194-195
Subiecți de cercetare, uman, protecția, 190-192
Respect pentru persoane, 191
Prejudecăți de răspuns, 109-111
Receptivitate, 165-167
Studii retrospective, 9, 105 Acid ribonucleic (ARN), 222 Factori de risc, 102
ARN (acid ribonucleic), 222
Curba ROC (caracteristică de funcționare a receptorului), 139
RR, vezi Risc relativ

Calculul mărimii eșantionului, 153-158
pentru diferența dintre două proporții, 158-159 pentru testarea diferenței dintre două medii, 159-160
Valori eșantion, 52-53 Metoda științifică, 1-17 Raționament științific, logica, 1-6 Screening, 129-139 Auto-raportare, 161 Sensibilitate, 130 Nivel de semnificație unic, 1-15
Studiu de semnificație unică, 1-15
polimorfisme de nucleotide (SNP), 180, 223-224 SMR (raportul de mortalitate standardizat), 94
SNP-uri (polimorfisme cu o singură nucleotidă), 180, 223-224
Formula Spearman-Brown, 163-164
Specificitate, 130 Abatere standard, 38 diferență între eroarea standard și, 42-45 sensul, 40-41
Eroare standard
diferența dintre abaterea standard și, 42-45 de diferență între două medii, 46-47 de medie, 43, 50 de proporție, 57 Rata mortalității standardizate (SMR), 94
distribuție normală standardizată, 47-51
Scor standard, 47-51 Metode statistice, 7-8
Puterea statistică, 154
Semnificație statistică, 153 Statistici, 6, 86
etica cercetării și, 189-196 Studii
design de, 8-10
tipuri de, 96-97
Citiri sugerate, 233-236 Analiza supraviețuirii, 117-120 Curbe de supraviețuire, 118-119 Timp de supraviețuire, 120

TDT (teste de dezechilibru de transmisie), 181-185
Teorie, 4-5
Teste de dezechilibru de transmisie (TDT), 181-185

t statistică, 51-52
valori critice ale, 199 *t* -test pentru diferența dintre două medii, 62-64 pereche potrivită, 66-67 performant, 64-66 utilizarea, 67-69
Studii pe gemeni, 173-175 ANOVA cu doi factori, 74-75 exemplu de, 76-77
Test cu două cozi, 64
test cu o singură coadă versus, 145-146 Eroare de tip I, 14, 16-17
consecințele, 16-17
Eroare de tip II, 14, 16-17
consecințele, 16-17

Valabilitate, 164-165
Variabilitate, 38
Variabile, 86
categoric, 10 confuz, 111-112 continuu, 10 dependent, 96 discret, 10 independent, 96 interacțiune între doi, 75-76
multiplicitate de, 167-168 selectând, pentru modele multivariate, 122-124 Varianta, 38, 163
analiză a, vezi Analiza variației între grupuri, 70 în interiorul grupurilor, 70
Diagramele Venn, 22

Scanarea întregului genom, 176-177
Testul sumelor de rang pentru perechi potrivite Wilcoxon, 67 Pătrat mediu în interiorul grupurilor, 72 Varianta în interiorul grupurilor, 70

Corecția și calculul lui Yates, 33

Scorul Z, 47-51
valori critice ale, 199
Z-test pentru compararea a două proporții, 60-62